

AC Q9M6A1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE GLYCINE-RICH RNA BINDING PROTEIN 1.
 GN GRP-1
 OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vinaceae; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Clastre M., Chenieux J.-C., Rideau M., Hamdi S.;
 RT "Genes encoding glycine-rich Catharanthus roseus proteins with RNA-
 binding motifs.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF200321; AAF31402.1; -;
 DR HSSP; P09651; 1HA1
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 SQ SEQUENCE 137 AA; 14162 MW; 4FADAB9C7A989FC CRC64;

Query Match 34.7%; Score 66; DB 10; Length 137;
 Best Local Similarity 50.0%; Pred. No. 1.8;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLROCLAAARAGGGGGGIEGP 26
 I : : : |||||
 Db 80 TVNEAQRSGGGGGGGRGP 101

RESULT 7

Q9M699
 ID Q9M699 PRELIMINARY; PRT; 160 AA.
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE GLYCINE-RICH RNA-BINDING PROTEIN 2.
 GN GRP-2.
 OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vinaceae; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Courtois M., Chenieux J.-C., Rideau M.,
 RA Clastre M.;
 RT "Cloning of two cDNAs encoding crGRP2 and crGRP3 (Accession Nos.
 RT AF200323 and AF200322), the first members of the RRM-GRP family in
 RT Catharanthus roseus (PGR00-049).";
 RL Plant Physiol. 122:1459-1459(2000).
 DR EMBL; AF200323; AAF31404.1; -;
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; rrm; 1.
 DR PRINTS; PR01228; EGGSELL.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 SQ SEQUENCE 160 AA; 16264 MW; DCDC9F63C983F5F2 CRC64;

Query Match 34.7%; Score 66; DB 10; Length 160;
 Best Local Similarity 50.0%; Pred. No. 2.1;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLROCLAAARAGGGGGGIEGP 26
 I : : : |||||
 Db 80 TVNEAQRSGGGGGGGRGP 101

RESULT 8

Q9XE89
 ID Q9XE89 PRELIMINARY; PRT; 369 AA.
 AC Q9XE89;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 39.1 KDA PROTEIN.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Liaca V., Lou A., Messing J.W.;
 RT "Microsytney analysis of 22-kDa zein cluster in maize and sorghum.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF061282; RAD22156.1; -;
 DR EMBL; AF061282; RAD22156.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 369 AA; 39080 MW; DAA3C65088F106CE CRC64;

Query Match 34.7%; Score 66; DB 10; Length 369;
 Best Local Similarity 40.0%; Pred. No. 4.5;
 Matches 14; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 4 PTLROCLAAARAGGGGGGIEPTLRO--CLAARA 36
 I : : : |||||
 Db 167 PAKKASIASVGGGGGGGVWRRRPGCGSGRS 201

RESULT 9

Q9LI26
 ID Q9LI26 PRELIMINARY; PRT; 413 AA.
 AC Q9LI26;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone: P0708G02.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP001539; BAA92912.1; -;
 DR EMBL; AP001539; BAA92912.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 413 AA; 45035 MW; 4FEC2A4C5D1271CF CRC64;

Query Match 34.7%; Score 66; DB 10; Length 413;
 Best Local Similarity 64.7%; Pred. No. 5;
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 ROCCLAAARAGGGGGGI 23
 I : : : |||||
 Db 162 RRCAGLLAGGGGGGV 178

RESULT 10

Q96SQ2
 ID Q96SQ2 PRELIMINARY; PRT; 474 AA.

34.7%; Query Match Score 66; DB 2; Length 496;
Best Local Similarity 46.2%; Pred. No. 6;
Matches 18; Conservative 2; Mismatches 9; Indels

PC TISSUE=BRAIN;
MEDLINE=21226021; PubMed=11326310;
RA Kato T., Satoh S., Okabe H., Kitahara O., Ono K., Kihara C.,
RA Tanaka T., Tsunoda T., Yamaoka Y., Nakamura Y., Furukawa Y.;
RT "Isolation of a novel human gene, MARKLI, homologous to MARK3 and its
RT involvement in hepatocellular carcinogenesis.";
RT Neoplasia 3:4-9(2001).
RL Neoplasia 3:4-9(2001).
CC EMBL; AB049127; BAB39380.1; -;
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR HSPD; Q63450; 1A06.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.

10 LAARA---GGGGGGGIEGPTLRQCLAA 34

RESULT 18	
99GNP1	PRT; 770 AA.
ID	PRELIMINARY;
Q9GNN1	
Q9GNP1;	
01-MAR-2001 (TReMBLrel. 16, Created)	
01-MAR-2001 (TReMBLrel. 16, Last sequence update)	
01-DSC-2001 (TReMBLrel. 19, Last annotation update)	
DE	VASA HOMOLOG.
CSDEAD1B(CSVHB).	
Clona savigny.	
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;	
Cnididae; Clona.	
NCBI_TaxID=51511;	
[1]	
SEQUENCE FROM N.A.	
TISSUE-Ovary:	
MEDLINE=20130953; PubMed=10664149;	
Fujimura M., Takamura K.;	
"Characterization of an ascidian DEAD-box gene, Ci-DEAD1: specific expression in the germ cells and its mRNA localization in the posterior-most blastomeres in early embryos.";	
Dev. Genes Evol. 210:64-72(2000).	
EMBL; AB047803; BAB12217.1;	
HSSP: Q58083; 1HV8.	
InterPro: IPR001410; DEAD.	
InterPro: IPR001650; Helicase_C.	
InterPro: IPR001878; znf_CCHC.	
Pfam: PF00270; DEAD; 1.	
Pfam: PF00271; helicase_C; 1.	
Pfam: PF00098; zf-CCHC; 6.	
SMART: SM00487; DEXDC; 1.	
SMART: SM00490; HELIC; 1.	
SMART: SM00343; znf_C2HC; 6.	
are-binding; Helicase; zinc-finger.	

```

SQ SEQUENCE 770 AA; 82032 MW; 5CD6D2A2D8C9CDD58 CRC64;
Query Match 34.5%; Score 65.5; DB 5; Length 770;
Best Local Similarity 43.2%; Pred No. 10;
Matches 16; Conservative 9; Mismatches 9; Indels 9; Gaps 2;

QY 2 EGPTLRQCLAAAGGGGGG-----GIEGPTLRQC 31
  || || : ||||| || || ||
Db 139 EGHNSREC-+PKGGGGGGGGCGFCGEGHMSREC 173
  || || : ||||| || || ||

RESULT 19
Q9LD54 PRELIMINARY; PRT; 381 AA.
ID Q9LD54
AC Q9LD54;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ESTS AU068633.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Poales; Poaceae;
OC Ehrhartoideae; Liliopsida; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone: p0453A06.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone: p0499C11.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC EMBL; AP001383; BAA92520.1; -.
DR EMBL; AP001080; BAA90348.1; -.
DR Interpro; IPR001993; Mitoch_carrier.
DR Interpro; IPR002067; Mit_carrier.
DR Pfam; PF00153; mito_carr; 3.
DR PRINTS; PS00926; MITOCARRIER.
DR PROSITE; PS00215; MITOCH_CARRIER; 2.
KW Inner membrane; Mitochondrion; Transmembrane; Transport.
SQ SEQUENCE 381 AA; 40761 MW; F3A0E3CED9D50778 CRC64;

Query Match 34.2%; Score 65; DB 10; Length 381;
Best Local Similarity 48.1%; Pred. No. 6.1;
Matches 13; Conservative 3; Mismatches 11; Indels 0; Gaps 0

QY 4 PTLRQCLAAAGGGGGGGGIEGPTLRQ 30
  || || : ||||| || || ||
Db 23 PHARRALALRVGGGGGGPAPASLTVRE 49
  || || : ||||| || || ||

RESULT 20
Q9VJK4 PRELIMINARY; PRT; 452 AA.
ID Q9VJK4
AC Q9VJK4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-NOV-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG5953 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
GN CG5953.

```


RA MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins A.G., Galie R.F.,
RA George R.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Brill J.F., Aquayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottiert P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glöckel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hariri N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Pacle J.M.,
RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svarksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003592; AAF51656.1; -.
DR GenBank; U07954; 1C8R.
DR FlyBase; FBgn0037034; CGI3257.
DR InterPro; IPR001545; Glyco_hormone_beta.
DR InterPro; IPR0021172; LDL_recept_A.
DR Pfam; PF00057; ldl_recept_a_1.
DR SMART; SM00192; LDLa; 1.
DR PROSITE; PS00261; GLYCO_HORMONE_BETA_1; UNKNOWN_1.
DR PROSITE; PS01209; LDLRa_1; 1.
DR PROSITE; PS00668; LDLRa_2; 1.
KW Glycoprotein.
SQ SEQUENCE 150 AA 15131 MW; 7A065C21C1376079 CRC64;
Query Match 33.9%; Score 64.5; DB 5; Length 150;
Best Local Similarity 51.6%; Pred.No.2.9;
Matches 16; Conservative 0; Mismatches 6; Indels 9; Gaps 1;
Qy 5 TLRQC-----AARAGGGGGGIEGP 26
|| || | | | | | | | | |
Db 43 TLLCLGLISLGMAATAAAAGGGGGGAAP 73
RESULT 22
Q9AF15 PRELIMINARY; PRT; 165 AA.
ID Q9AF15
AC Q9AF15;
DT 01-JUN-2001 [TrEMBLrel. 17, Created]
DT 01-JUN-2001 [TrEMBLrel. 17, Last sequence update]
DE 01-DEC-2001 [TrEMBLrel. 19, Last annotation update]
DE SINGLE-STRANDED DNA-BINDING PROTEIN.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxId=1772;

[1]
SEQUENCE FROM N.A.
Reddy M.S., Muniyappa K.:
"Biochemical properties of single-stranded DNA-binding proteins from
Mycobacteria."
Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF349434; AAK30583.1; -
DR HSSP: P02339; 1EYG.
DR InterPro: IPR000424; SSB.
DR Pfam: PF00436; SSB; 1.
DR
SQ
SEQUENCE 165 AA; 17401 MW; 8786415C16F26F39 CRC64;
Query Match 33.9%; Score 64.5; DB 2; Length 165;
Best Local Similarity 48.5%; Pred. No. 3.1;
Matches 16; Conservative 9; Indels 5; Gaps 1;
QY 3 GPTLROCL-----AARAGGGGGGGIEGPTLRQ 30
||:|||||:|||||||:|:|
DB 107 GPSLRVATKVNKASRGGGGGFGSGGGSQR 139
|:|:|:|:|:|:|:|:|:|:|
RESULT 23
Q9VW01 PRELIMINARY: PRT; 309 AA.
AC Q9VW01;
Q9VW01;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG13055 PROTEIN.
DE CG13055.
GN Drosophila melanogaster (Fruit fly).
OS
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
OX [1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RP MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu P.V., Beran B.P., Bock J., Brockstein P., Brottier P.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bouck J., Cadieu E., Center A., Chandra I.,
RA Borkova D., Botchan M.R., Boutler H., Cadenhead L.B., Davies P.,
RA Burtis K.C., Busam D.A., Butler H., Cadenhead L.B., Davies P.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Wei M.-H., Ibegwam C.,
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Li X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Morbarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S.-Y., Yao Q.A.,

Query Match 33.48; Score 63.5; DB 10; Length 113;
Best Local Similarity 36.18; Pred. No. 2.8;
Matches 13; Conservative 6; Mismatches 7; Gaps 1;

8 QCLAAR-----AGGGGGGGIEGPTLRQCLAARA 36 QY
: : : | | | | | : | | : :
71 KCMCTKRCGGGGGGGGGGGGDEPPLREARVHRS 106 DB

Search completed: October 9, 2002, 09:03:09
Job time : 13.9826 secs

3312 VPGVWVLRCPDAGAGGGGGG 3333

RESULT 29

Q9S0R8 PRT: 3972 AA.

ID	Q950R8;
AC	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DT	01-DEC-2001 POLYMERASE SYNTHASE AVES 1.

TYPE 1 POLYMERASE CHAIN REACTION
DE AMPLIFICATION OF DNA FROM
GN AVEAL.
OS Streptomyces avermitilis.
SC Bacteria; Firmicutes; Actinobacteria;
OC Bacteriota; Actinomycetales; Streptomycetaceae; Streptomyces.
CX Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=33903;
[1]
SEQUENCE FROM N.A.

SEQUENCE: 3972 AA; 416852 MW. 2A293695B032B1C3 CRC64;
 phosphopantetheine; transference.
 PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
 PROSITE; PS00133; CARBOXYPEPT_2N_2; UNKNOWN_2.
 PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 PROSITE; PS50075; ACP_DOMAIN; 3.
 Pfam; PF00550; pp-binding; 3.
 Pfam; PF02801; ketoacyl-synt_C; 2.
 Pfam; PF00109; ketoacyl-synt; 2.
 Pfam; PF00698; Acyl_transf; 3.
 InterPro; IPR000834; zn_carboOept.
 InterPro; IPR003880; Phosphopant_attach.
 InterPro; IPR000794; ketoacyl-synt.
 InterPro; IPR001227; Acyltransfer_domain.
 EMBL; BA084474.1; -
 PROSITE; PS00367; BAA84474.1; -
 "Organization of the biosynthetic gene cluster for the polyketide
 avermectin in *Streptomyces avermitilis*.";
 Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514 (1999).
 Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
 MEDLINE=99380548; PubMed=10449723;
 SEQUENCE: 39380548;

Query Match 33.7%; Score 64; DB 2; Length 3972;
Best Local Similarity 54.5%; Pred. No. 73;
Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLARAGGGGGG 22
: | : | | | | | | | |
db 3345 VPGVVLRCPDAGAGGGGGG 3366

RESULT 30

Q942U6
PRELIMINARY:
PRT; 113 AA.

Q94206	FREEMANIN
ID	
AC	Q94206; 19, Created)
AD	(TREMUREL
DT	01-DEC-2001
DT	01-DEC-2001
DT	01-DEC-2001
DT	01-DEC-2001
DE	P0506E04.26
DE	PROTEIN.
DN	P0506E04.26
OS	Oryza sativa (Rice).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC	Ehrhartoidae; Oryzeae; Oryza.
OX	NCBI_TaxID=4530;
OX	

[1] SEQUENCE FROM N.A.
 RN STRAIN-CV, NIPPONBARE;
 RC STRAIN-T, Matsumoto T., Yamamoto K.;
 RA "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 1, PAC
 RT clone:P0506E04." -;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RI EMBL: AP003772; BAB67948.1; -;
 DR EMBL: AP003772; BAB67948.1; -;
 DB EMBL: AP003772; BAB67948.1; -;
 SO SEQUENCE 113 AA; 11708 MW; 26D9B2C86935BC0B CRC64;

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	190	100.0		36	21	AAB17298	TPO-mimetic peptid
2	190	100.0		36	21	AAB17299	TPO-mimetic peptid
3	190	100.0		36	21	AAV96521	Cyclic or linear t
4	172	90.5		36	21	AAB17300	TPO-mimetic peptid
5	172	90.5		36	21	AAV96522	Linear thrombopoie
6	168	88.4		36	21	AAB16963	TPO-mimetic peptid
7	168	88.4		36	21	AAB17293	TPO-mimetic peptid
8	168	88.4		36	21	AAV96525	Thrombopoietin mim
9	168	88.4		41	21	AAV96528	Thrombopoietin mim
10	168	88.4		42	21	AAB17281	TPO-mimetic peptid
11	168	88.4		42	21	AAB17282	TPO-mimetic peptid

DR WPI; 2000-350702/30.

Wed Oct 9 10:29:54 2002

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
XX Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 36 AA;

Query Match 100.0%; Score 190; DB 21; Length 36;
Best Local Similarity 100.0%; Pred. No. 3.8e-16; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 0;

QY 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
|||||
Db 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36

RESULT 3
AAB17299
ID AAB17299 standard; Peptide: 36 AA.

XX AAB17299;
XX
XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:355.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

XX Synthetic.
XX
XX WO2000024782-A2.
XX
XX 04-MAY-2000.
XX
XX 25-OCT-1999; 99WO-US25044.
XX
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX
XX (AMGE-) AMGEN INC.
XX
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI; 2000-350702/30.
XX Novel composition of matter comprising an Fc domain and
PT

PT pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -
XX Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 36 AA;

Query Match 100.0%; Score 190; DB 21; Length 36;
Best Local Similarity 100.0%; Pred. No. 3.8e-16; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 0;

QY 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
|||||
Db 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36

RESULT 3
AAB17299
ID AAB17299 standard; Peptide: 36 AA.

XX AAB17299;
XX
XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:355.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

XX Synthetic.
XX
XX WO2000024782-A2.
XX
XX 04-MAY-2000.
XX
XX 25-OCT-1999; 99WO-US25044.
XX
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX
XX (AMGE-) AMGEN INC.
XX
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI; 2000-350702/30.
XX Novel composition of matter comprising an Fc domain and
PT

PI Liu C, Feige U, Cheetham J;
 DR WPI; 2000-365108/31.
 XX
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 XX
 PS Claim 16; Page 61; 9lpp; English.
 XX
 CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP-1-(L1)-TMP-2),
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₁-X_{1,0}, X₂-X_{1,1}, X₂-X_{1,2},
 CC X₂-X_{1,3}, X₂-X_{1,4}, X₁-X_{1,0}, X₁-X_{1,1}, X₁-X_{1,2}, X₁-X_{1,3}, and
 CC X₁-X_{1,4}. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X_{1,0} = L, I, V, A, F, M, or G; X_{1,1} = A, I, V,
 CC L, F, S, T, K, H, or E; X_{1,2} = A, I, V, L, F, G, S, or Q; X_{1,3} = R, K,
 CC T, V, N, Q or G; X_{1,4} = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 XX
 SQ Sequence 36 AA;
 Query Match 100.0%; Score 190; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 3.8e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 Db 1 IEPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 RESULT 4
 AABI7300
 ID AABI7300 standard; Peptide; 36 AA.
 AC AABI7300;
 XX
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:356.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 FT W0200024782-A2.
 XX
 FN 04-MAY-2000.
 XX
 PD 25-OCT-1999; 99WO-US25044.
 XX
 PF 23-OCT-1998; 98US-0105371.
 XX
 PR 22-OCT-1999; 99US-0428082.
 XX
 XX (AMGE-) AMGEN INC.
 PA Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.
 DR

XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX Example 1; Page 321; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P*3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation and possibly placental transfer. AAG69443
 CC to AAG69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX
 SQ Sequence 36 AA;
 Query Match 90.5%; Score 172; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 5.5e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 Db 1 IEPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 RESULT 5
 AAY96522
 ID AAY96522 standard; peptide; 36 AA.
 XX
 AC AAY96522;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Linear thrombopoietin mimetic peptide compound 3.
 XX
 KW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; linear.
 XX
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT 1..14
 FT /label= TMP_1
 FT Peptide 15..22
 FT /label= linker
 FT Peptide 23..36
 FT /label= TMP_2
 XX
 PN W0200024770-A2.
 XX
 XX 04-MAY-2000.
 XX
 XX 22-OCT-1999; 99WO-US24834.
 XX
 XX 23-OCT-1998; 98US-0105348.
 XX
 XX (AMGE-) AMGEN INC.
 XX

Wed Oct 9 10:29:54 2002

PI Liu C, Feige U, Cheetham J;
 XX WPI: 2000-365108/31.
 DR
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Claim 16; Page 61; 91pp; English.
 XX
 CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP₁-(L₁)-TMP₂],
 CC is new. TMP₁ and TMP₂ are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂,
 CC X₂-X₁-L₃, X₂-X₁-L₄, X₁-X₁-L₀, X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and
 CC X₁-X₁-L₄. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A, F, M, or K; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, T, R, E, or G; X₁₃ = R, K,
 CC T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 XX Sequence 36 AA;
 SQ
 Query Match 90.5%; Score 172; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 5.5e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEGPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 DB 1 IEGPTLRQALAAARAGGGGGGIEGPTLRQALAAARA 36
 RESULT 6
 AAB16963
 ID AAB16963 standard; Protein: 36 AA.
 XX
 AC AAB16963;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide TMP-TMP SEQ ID NO:14.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI: 2000-350702/30.
 DR
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Disclosure; Page 190; 608pp; English.
 XX
 CC The present invention describes composition of matter (1) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (1) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAB69443
 CC to AAB69526 and AAB16935 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX Sequence 36 AA;
 SQ
 Query Match 88.4%; Score 168; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 1.6e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEGPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 DB 1 IEGPTLRQALAAARAGGGGGGIEGPTLRQALAAARA 36
 RESULT 7
 AAB17293
 ID AAB17293 standard; Peptide: 36 AA.
 XX
 AC AAB17293;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:349.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI: 2000-350702/30.
 DR
 PT Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -
PS
XX Example 1; Page 318; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antitumoral, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAG69443 to AAG69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Sequence 36 AA;

Query Match 88.4%; Score 168; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 1.6e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
IIIIIIII IIIIIIIIIIIIIIIIIIIII IIIIIII
Db 1 IEPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 8

AAAY96525
ID AAAY96525 standard; peptide; 36 AA.

AC AAAY96525;

DT 04-SEP-2000 (first entry)

Thrombopoietin mimetic peptide compound 6.

Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
immunosuppressive; anti-inflammatory; linker.
Synthetic.

Key	Location/Qualifiers
Modified-site	1
Peptide	/note= "optionally linked to an Fc molecule"
Peptide	1..14
Peptide	/label= TMP_1
Peptide	15..18
Peptide	/label= linker
Peptide	19..32
Modified-site	/label= TMP_2
Modified-site	32
Modified-site	/note= "optionally linked to an Fc molecule"

WO200024770-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24834.

23-OCT-1998; 98US-0105348.

(AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;
XX WPI; 2000-365108/31.
XX

Thrombopoietic peptides which activate mpl receptors and increase the production of platelets or platelet precursors, useful for treatment of diseases which involve thrombocytopenia
Claim 16; Page 62; 91pp; English.

A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L1)-TMP_2], is new. TMP_1 and TMP_2 are amino acid sequences varying from at least 10 to 14 residues in length comprising X₂-X₁-X₀, X₂-X₁-X₁, X₂-X₁-X₁-X₂, X₂-X₁-X₁-X₁-X₂, X₁-X₁-X₁-X₁, X₁-X₁-X₁-X₁-X₂, and X₁-X₁-X₁-X₁-X₁-X₂, where X₀ = L, I, V, A, F, M, or K; X₁ = R or Q; X₂ = G or A; X₃ = P; X₄ = T or S; X₅ = T or S; X₆ = L, I, V, A, F, M, or K; X₇ = R or Q; X₈ = Q, N, or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, L, F, S, T, R, E, or G; X₁₂ = A, I, V, L, F, S, T, R, E, or G; X₁₃ = A, I, V, L, F, S, T, R, E, or G; X₁₄ = A, I, V, L, F, S, T, R, E, or G; X₁₅ = A, I, V, L, F, S, T, R, E, or G; X₁₆ = A, I, V, L, F, S, T, R, E, or G; X₁₇ = A, I, V, L, F, S, T, R, E, or G; X₁₈ = A, I, V, L, F, S, T, R, E, or G; X₁₉ = A, I, V, L, F, S, T, R, E, or G; X₂₀ = A, I, V, L, F, S, T, R, E, or G. The compounds bind to and activate the c-Mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

Sequence 36 AA;

Query Match 88.4%; Score 168; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 1.6e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
IIIIIIII IIIIIIIIIIIIIIIIIIIII IIIIIII
Db 1 IEPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 9

AAAY96528

ID AAAY96528 standard; peptide; 41 AA.

AC AAAY96528;

DT 04-SEP-2000 (first entry)

Thrombopoietin mimetic peptide compound 9.

Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
immunosuppressive; anti-inflammatory; linker.
Synthetic.

Key	Location/Qualifiers
Modified-site	1
Peptide	/note= "optionally linked to an Fc molecule"
Peptide	6..19
Peptide	/label= TMP_1
Peptide	20..27
Peptide	/label= linker
Peptide	28..41
Peptide	/label= TMP_2

WO200024770-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24834.

23-OCT-1998; 98US-0105348.

us-09-422-838c-28.ra.g

Wed Oct 9 10:29:54 2002

PA (AMGE-) AMGEN INC.
 PI Liu C, Feige U, Cheetham J;
 XX WPI; 2000-365108/31.
 DR
 XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Claim 16; Page 65; 91pp; English.
 XX
 CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP-1-(L1-L1)-TMP-2),
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₁-X_{1,0}, X₂-X_{1,1}, X₂-X_{1,2},
 CC X₂-X_{1,3}, X₂-X_{1,4}, X₁-X_{1,0}, X₁-X_{1,1}, X₁-X_{1,2}, X₁-X_{1,3}, and
 CC X₁-X_{1,4}. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A, F, M, or K; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, S, or Q; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, T, R, E, or G; L₁ = linker.
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, e.g.
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 SQ Sequence 41 AA;
 Query Match 88.4%; Score 168; DB 21; Length 41;
 Best Local Similarity 94.4%; Pred. No. 1.9e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 36
 DB 6 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 41
 RESULT 10
 AAB17281
 ID AAB17281 standard; Peptide; 42 AA.
 AC AAB17281;
 XX
 XX 31-OCT-2000 (first entry)
 DE TPO-mimetic peptide sequence SEQ ID NO:337.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CIL44; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX Synthetic.
 OS WO200024782-A2.
 XX
 XX 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.
 XX
 XX WPI; 2000-350702/30.
 XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Disclosure; Page 313; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present inventions are
 CC be used for producing cancer, asthma, thrombosis, or autoimmune diseases.
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAB69443
 CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 42 AA;
 Query Match 88.4%; Score 168; DB 21; Length 42;
 Best Local Similarity 94.4%; Pred. No. 1.9e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 36
 DB 7 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 42
 RESULT 11
 AAB17282
 ID AAB17282 standard; Peptide; 42 AA.
 AC AAB17282;
 XX
 XX 31-OCT-2000 (first entry)
 DE TPO-mimetic peptide sequence SEQ ID NO:338.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CIL44; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX Synthetic.
 OS WO200024782-A2.
 XX
 XX 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.
 XX

CC	Overlapping oligonucleotides were used to construct a synthetic
CC	gene encoding a thrombopoietin mimetic peptide (TMP), which
CC	was fused in-frame to the FC region of the human IgG1 chain (see
CC	AAY96529). A compound which binds to an mpl receptor comprising a TMP
CC	dimer joined by a linker [TMP_1-(L_1)-nTMP_2], is new. TMP_1 and TMP_2
CC	are amino acid sequences varying from at least 10 to 14 residues in
CC	length comprising X ₂ -X ₁ -0, X ₂ -X ₁ -1, X ₂ -X ₁ -2, X ₂ -X ₁ -3, X ₂ -X ₁ -4,
CC	X ₁ -X ₁ -0, X ₁ -X ₁ -1, X ₁ -X ₁ -2, X ₁ -X ₁ -3, and X ₁ -X ₁ -4. X ₁ = I, A,
CC	V, L, S or R; X ₂ = E, D, K or V; X ₃ = G or A; X ₄ = P; X ₅ = T or S;
CC	X ₆ = L, I, V, A or F; X ₇ = R or K; X ₈ = Q, N, or E; X ₉ = W, Y or F;
CC	X ₁₀ = L, I, V, A, F, M, or Q; X ₁₁ = A, I, V, L, F, S, T, K, H, or E;
CC	X ₁₂ = A, I, V, L, F, G, S, or Q; X ₁₃ = R, K, T, V, N, O or G; X ₁₄ =
CC	A, I, V, L, T, R, E, or G; L ₁ = linker comprising 1 to 20 amino
CC	acids; and n = 0 or 1. The compounds bind to and activate the c-mpl
CC	receptor which mediates the activity of endogenous thrombopoietin. The
CC	TMPs are useful for increasing the production of platelets or platelet
CC	precursors (e.g. megakaryocytes) in a mammal, which is useful for
CC	treatment of diseases which involve thrombocytopenia, e.g. aplastic
CC	anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
CC	associated ITP, and systemic lupus erythematosus.
XX	
XX	
SQ	Sequence 42 AA;
	Query Match 88.4%; Score 168; DB 21; Length 42;
	Best Local Similarity 94.4%; Pred. No. 1.9e-13;
	Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 IEPTTLRQCIAARAGGGGGIEGPTLRQCLAAARA 36
DB	7 IEPTLRQWLAAARAGGGGGIEGPTLRQLWLAARA 42
RESULT 14	
ID	AAB17311 standard; Peptide; 60 AA.
XX	
AC	AAB17311;
DT	
XX	31-OCT-2000 (first entry)
DE	Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.
XX	
KW	Modified peptide; therapeutic agent; fusion; FC domain; cancer;
KW	autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW	immunosuppressive; epo; tpo; CTU44; mimetic; IL-1; TNF; antagonist;
KW	MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW	vascular endothelial growth factor; matrix metalloproteinase;
KW	asthma; thrombosis; pharmaceutical.
XX	
OS	Homo sapiens.
OS	Synthetic.
PN	WO200024782-A2.
XX	
PD	04-MAY-2000.
XX	
PF	25-OCT-1999; 99WO-US25044.
XX	
PR	23-OCT-1998; 98US-0105371.
XX	
PR	22-OCT-1999; 99US-0428082.
XX	
PA	(AMGE-) AMGEN INC.
PI	Feige U, Liu C, Cheetham J, Boone TC;
DR	WPI; 2000-350702/30.
XX	
PT	Novel composition of matter comprising an Fc domain and
PT	pharmacologically active peptides, useful for treating cancer and
PT	autoimmune diseases -
PS	Example 2; Page 331; 608pp; English.

X	The present invention describes composition of matter (I) comprising an
XX	Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XC	(X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC	independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC	-(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC	where P1, P2, P3, and P4 = are each independently sequences of
CC	pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC	independently linkers; and a, b, c, d, e, and f = are each independently
CC	have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC	activities. DNAs, vectors and host cells from the present invention can
CC	be used for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC	The use of an Fc domain (rather than a Fab domain) can provide a longer
CC	half-life or incorporate functions such as Fc receptor binding, protein
CC	A binding, complement fixation, and possibly placental transfer. AAA69443
CC	A binding, complement fixation, and possibly placental transfer. AAA69443
CC	sequences used in the exemplification of the present invention.
CC	
XX	Sequence 60 AA:
SQ	
Query Match	88.4%; Score 168; DB 21; Length 60;
Best Local Similarity	94.4%; Pred. No. 2.7e-13;
Matches 34; Conservative	0; Mismatches 2; Indels 0; Gaps 0
QY	1 IEPTTLRQLAARAGGGGGGIEGPTLRQCLAAARA 36
DB	
	2 IEPTTLRQLAARAGGGGGGIEGPTLRQCLAAARA 37
RESULT 15	
AAAB16960	
ID	AAB16960 standard; Protein: 269 AA.
AC	
AC	AAAB16960;
XX	
DT	31-OCT-2000 (first entry)
XX	
DE	TMP-TWP-Fc protein sequence SEQ ID NO:10.
XX	
KW	Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
KW	autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW	immunosuppressive; Epo; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW	MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW	vascular endothelial growth factor; matrix metalloproteinase;
KW	asthma; thrombosis; pharmaceutical.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PN	WO200024782-A2.
XX	
PD	04-MAY-2000.
XX	
Pf	25-OCT-1999; 99WO-US255044.
XX	
PR	23-OCT-1998; 98US-0105371.
PR	22-OCT-1999; 99US-0428082.
XX	
PA	(AMGE-) AMGEN INC.
XX	
PI	Felge U, Liu C, Cheetham J, Boone TC;
XX	
XX	WPI: 2000-350702/30.
DR	N-PSDB; AAA69446.
DR	
XX	
PT	Novel composition of matter comprising an Fc domain and
PT	pharmacologically active peptides, useful for treating cancer and
PT	autoimmune diseases -
XX	
PS	Example 2: Page 185-186; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA;

Query Match 88.4%; Score 168; DB 21; Length 269;
 Best Local Similarity 94.4%; Pred. No. 1.2e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 36
 DB 2 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 37

RESULT 16

AAAY96531
 ID AAY96531 standard; Protein; 269 AA.

AC AAY96531;

DT 04-SEP-2000 (first entry)

DE Human IgG1 Fc TWP fusion protein.

XX Immunoglobulin; IgG1; Fc: thrombopoietin; mimetic; TWP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX Homo sapiens.

OS WO200024770-A2.

PN 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

DR N-PSDB; AAA29229.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 CC production of platelets or platelet precursors, useful for treatment of
 CC diseases which involve thrombocytopenia

PS Example 2a; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TWP) dimer joined by a linker [TWP-1-(L1)-TWP-2],
 CC is new. TWP-1 and TWP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X2-X1_0, X2-X1_1, X2-X1_2,
 CC X2-X1_3, X2-X1_4, X1-X1_0, X1-X1_1, X1-X1_2, X1-X1_3, and

CC X1-X1_4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
 CC or A; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
 CC or E; X9 = W, Y or F; X10 = L, I, V, A, F, W, or K; X11 = A, I, V,
 CC L, F, S, T, K, H, or E; X12 = A, I, V, L, F, G, S, or Q; X13 = R, K,
 CC T, V, N, Q or G; X14 = A, I, V, L, F, T, R, E, or G; L1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TWPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 269 AA;

Query Match 88.4%; Score 168; DB 21; Length 269;
 Best Local Similarity 94.4%; Pred. No. 1.2e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 36
 DB 234 IEGPTLRQCLARAGGGGGGIEGPTLRQCLARA 269

RESULT 17

AAB16959

ID AAB16959 standard; Protein; 268 AA.

AC AAB16959;

DT 31-OCT-2000 (first entry)

DE Fc-TMP-TMP protein sequence SEQ ID NO:8.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.

OS Synthetic.

PN WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

DR N-PSDB; AAA69445.

XX Novel composition of matter comprising an Fc domain and
 CC pharmacologically active peptides, useful for treating cancer and
 CC autoimmune diseases

PS Example 2; Page 182-183; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of

Wed Oct 9 10:29:54 2002

0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Query Match 84.2%; Score 160; DB 21; Length 36;
Best Local Similarity 91.7%; Pred. No. 1.5e-12;
Matches 33; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 36
||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 36

RESULT 19
AAY96523
ID AAY96523 standard; peptide; 36 AA.
XX
AC AAY96523;
XX
DT 04-SEP-2000 (first entry)
XX
DE Thrombopoietin mimetic peptide compound 4.
XX
KW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "optionally linked to an Fc molecule"
FT Peptide 1.14 /label= TMP_1
FT Peptide 15..22 /label= linker
FT Modified-site 18
FT Peptide 23..36 /note= "optionally modified by bromoacetyl or PEG"
FT Peptide /label= TMP_2
XX
PN WO200024770-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24834.
XX
PR 23-OCT-1998; 98US-0105348.
XX
PA (AMGE-) AMGEN INC.
XX
PI Liu C, Feige U, Cheetham J;
XX
DR WPI; 2000-365108/31.
XX
PS Thrombopoietic peptides which activate mpl receptors and increase the production of platelets or platelet precursors, useful for treatment of diseases which involve thrombocytopenia
XX
PS Claim 16; Page 62; 91pp; English.
XX
CC A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L1)-TMP_2],
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Query Match 86.3%; Score 164; DB 21; Length 268;
Best Local Similarity 94.3%; Pred. No. 3.6e-12;
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 35
||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 234 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 268

RESULT 18
AAB17301
ID AAB17301 standard; Peptide; 36 AA.
XX
AC AAB17301;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:357.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CMHA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
XX
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PS Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -
XX
PS Example 1; Page 321; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 37 AA;

Query Match 82.9%; Score 157.5; DB 21; Length 37;
 Best Local Similarity 91.9%; Pred. No. 3.1e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 Oy 1 IEGPTLRQCLAAARA--GGGGGGGIEGPTLRQCLAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEGPTLRQWLAAARAGGGGGGGGIEGPTLRQWLAAARA 37

RESULT 24
 AAB17295
 ID AAB17295 standard; Peptide; 38 AA.
 XX
 AC AAB17295;
 XX

31-OCT-2000 (first entry)
 XX

TPO-mimetic peptide sequence SEQ ID NO:351.
 XX

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX

Synthetic.
 OS
 XX

WO2000024782-A2.
 PN
 XX

04-MAY-2000.
 PD
 XX

25-OCT-1999; 99WO-US25044.
 PF
 XX

23-OCT-1998; 98US-0105371.
 PR
 XX

22-OCT-1999; 99US-0428082.
 PR
 XX

(AMGE-) AMGEN INC.
 PA
 XX

Feige U, Liu C, Cheetham J, Boone TC;
 PI
 XX

WPI; 2000-350702/30.
 DR
 XX

Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 PT
 XX

Example 1; Page 319; 608pp; English.
 PS
 XX

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX

SQ Sequence 38 AA;

Query Match 82.6%; Score 157; DB 21; Length 38;
 Best Local Similarity 89.5%; Pred. No. 3.6e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 2; Gaps 1;
 Oy 1 IEGPTLRQCLAAARA--GGGGGGGIEGPTLRQCLAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEGPTLRQWLAAARAGGGGGGGGIEGPTLRQWLAAARA 38

RESULT 25

AAB17304

ID AAB17304 standard; Peptide; 39 AA.

XX

AC AAB17304;

XX

31-OCT-2000 (first entry)

XX

TPO-mimetic peptide sequence SEQ ID NO:360.

XX

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

Synthetic.

OS

WO2000024782-A2.

XX

04-MAY-2000.

XX

25-OCT-1999; 99WO-US25044.

XX

23-OCT-1998; 98US-0105371.

PR

22-OCT-1999; 99US-0428082.

XX

(AMGE-) AMGEN INC.

PA

XX

Feige U, Liu C, Cheetham J, Boone TC;

PI

XX

WPI; 2000-350702/30.

DR

XX

Novel composition of matter comprising an Fc domain and

pharmacologically active peptides, useful for treating cancer and

autoimmune diseases -

PT

XX

Example 1; Page 323; 608pp; English.

PS

XX

CC The present invention describes composition of matter (I) comprising an

Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

(X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

-(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

where P1, P2, P3, and P4 = are each independently sequences of

pharmacologically active peptides; L1, L2, L3, and L4 = are each

independently linkers; and a, b, c, d, e, and f = are each independently

0 or 1, provided that at least 1 of a and b is 1. The composition can

have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP₁-(L₁L)_n TMP₂], where n is 1 to 14, L₁ and TMP₂ are amino acid sequences varying from at least 1 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂, X₂-X₁-L₃, X₂-X₁-L₄, X₁-L, V, L, S or R; X₂=E, D, K or V; X₃=G or A; X₄=P; X₅=T or S; X₆=L, I, V, A, F, M, or K; X₇=R or K; X₈=Q, N, V, S, T, K, H, or E; X₁₁-L₂=A, I, V, L, F, G, S, or O; X₁₁-L₁=A, I, V, F, S, T, K, H, or E; X₁₁-L₀=A, I, V, L, F, G, S, or O; X₁₃=L, V, K, Q or G; X₁₄=A, I, V, L, T, R, E, or F; L₁=linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid identification of the present invention.

CC sequences used in the exemplification of the procedure are:

XX	
Sequence	42 AA;
SQ	91 58; Score 155; DB 21; Length 42;

[illegible]

Search completed: October 9, 2002, 08:58:57
Job time : 17.1874 secs

RESULT 30

AAB17292 standard: Peptide; 35 AA.

XX
ID AABT/232

AC AAB17292;

XX
RT
31-OCT-2000 (first entry)

DI
XX

01 OCT 2009 (17:26) ...
... cont'do sequence SEQ ID NO: 348.

XX Modified peptide; therapeutic agent; fusion; FC domain; cancer;
KW autoimmunity; disease; cytostatic; antitumoral; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonistic;
KW MMP; inhibitory; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic; T cell lymphocyte antigen 4; tumour necrosis factor;
KW endothelial growth factor; matrix metalloproteinase;
KW vascular thrombolysis; pharmaceutical.

AA Synthetic.

XX
PN W0200024782-A2.

PN
XX
PD

FD 04 MAR 2001
XX
PF 25-OCT-1999; 99WO-US25044.

23 OCT 1998
XX
23-OCT-1998: 98US-0105371.

22-OCT-1999; 99US-0428082.

XX
PA (AMGE-) AMGEN INC.

FA	(range)	FA	Boone TC:
XX	10-15	Cheetham J.	
XX	15-20	Cheetham J.	
XX	20-25	Cheetham J.	
XX	25-30	Cheetham J.	
XX	30-35	Cheetham J.	
XX	35-40	Cheetham J.	
XX	40-45	Cheetham J.	
XX	45-50	Cheetham J.	
XX	50-55	Cheetham J.	
XX	55-60	Cheetham J.	
XX	60-65	Cheetham J.	
XX	65-70	Cheetham J.	
XX	70-75	Cheetham J.	
XX	75-80	Cheetham J.	
XX	80-85	Cheetham J.	
XX	85-90	Cheetham J.	
XX	90-95	Cheetham J.	
XX	95-100	Cheetham J.	

PI Felge U, Liu C, Chen
XX
DB WPT: 2000-350702/30.

Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

XX

Example 1, page 317, col. 1, line 10-11, should read:

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers, where (I) is: (x1)a-f1-(x2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4-(L1)c-p1-(L2)d-p2-(L3)e-p3, or -(L1)c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4 where F1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA6944

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 ; Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-28

Perfect score: 190

Sequence: 1 IEPTLRQCLAAARAGGGGGGIEPTLRQCLAAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.5	33.9	991	4	US-09-352-159-27
2	64.5	33.9	991	4	US-09-352-159-27
3	64.5	33.9	1196	4	US-09-352-159-31
4	64.5	33.9	1196	4	US-09-352-159-31
5	64	33.7	584	2	US-08-987-466-4
6	64	33.7	584	2	US-09-240-359-4
7	62	32.6	440	3	US-09-100-664A-2
8	62	32.6	440	3	US-09-100-664A-3
9	62	32.6	440	3	US-09-100-664A-3
10	60	31.6	14	2	US-08-764-640-13
11	60	31.6	14	2	US-08-764-640-13
12	60	31.6	14	2	US-08-764-640-13
13	60	31.6	14	2	US-08-764-640-13
14	60	31.6	14	2	US-08-764-640-13
15	60	31.6	14	2	US-08-764-640-13
16	60	31.6	14	2	US-08-764-640-13
17	60	31.6	14	2	US-08-764-640-13
18	60	31.6	14	2	US-08-764-640-13
19	60	31.6	14	2	US-08-764-640-13
20	60	31.6	14	2	US-08-764-640-13
21	60	31.6	14	2	US-08-764-640-13
22	60	31.6	14	2	US-08-764-640-13
23	60	31.6	14	2	US-08-764-640-13
24	60	31.6	14	2	US-08-764-640-13
25	60	31.6	14	2	US-08-764-640-13
26	60	31.6	14	2	US-08-764-640-13
27	60	31.6	14	2	US-08-764-640-13

28	60	31.6	16	2	US-08-764-640-232
29	60	31.6	16	3	US-08-973-225-18
30	60	31.6	16	3	US-08-973-225-194
31	60	31.6	16	3	US-08-973-225-194
32	60	31.6	16	3	US-08-973-225-220
33	60	31.6	16	3	US-09-244-298A-18
34	60	31.6	16	3	US-09-244-298A-194
35	60	31.6	16	3	US-09-244-298A-232
36	60	31.6	16	4	US-09-516-704-18
37	60	31.6	16	4	US-09-516-704-194
38	59.5	31.3	26	1	US-09-516-704-232
39	59.5	31.3	126	1	US-07-776-272-16
40	59.5	31.3	126	2	US-08-451-947-57
41	59.5	31.3	126	3	US-08-424-826A-57
42	59.5	31.3	126	5	US-08-928-694-57
43	59.5	31.3	969	2	PCT-US91-06950-57
44	59.5	31.3	969	2	US-08-284-941-2
45	59.5	31.3	969	4	US-08-447-642-2
					US-09-236-503-2

ALIGNMENTS

RESULT 1
US-09-352-159-27
; Sequence 27, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 991
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-159-27

Query Match 33.9%; Score 64.5; DB 4; Length 991;
Best Local Similarity 54.2%; Pred. No. 5.1;
Matches 13; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 3 GPTLRQCLAAARAGGGGGGIEG 25
|||: | | | | | | | | | |
Db 503 GFSIPPCADGAKAGGGGGSGG 526

RESULT 2
US-09-352-168-27
; Sequence 27, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875

Sequence 232, App
Sequence 18, Appl
Sequence 194, App
Sequence 220, App
Sequence 18, Appl
Sequence 194, App
Sequence 232, App
Sequence 18, Appl
Sequence 194, App
Sequence 232, App
Sequence 16, Appl
Sequence 57, Appl
Sequence 57, Appl
Sequence 57, Appl
Sequence 57, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

Sequence 232, App
Sequence 18, Appl
Sequence 194, App
Sequence 220, App
Sequence 18, Appl
Sequence 194, App
Sequence 232, App
Sequence 18, Appl
Sequence 194, App
Sequence 232, App
Sequence 16, Appl
Sequence 57, Appl
Sequence 57, Appl
Sequence 57, Appl
Sequence 57, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

us-09-422-838c-28.ra1

Wed Oct 9 10:29:55 2002

```
;
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 991
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-168-27

Query Match          33.9%; Score 64.5; DB 4; Length 991;
Best Local Similarity 54.2%; Pred. No. 5.1;
Matches 13; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 3 GPTLRQCL-AARAGGGGGGGIEG 25
   ||:| | | | | | | | | | |
Db 503 GPSIPPCADGAKAGGGGGGGSGG 526

RESULT 3
US-09-352-159-31
; Sequence 31, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Maddox, Joyce R.
; APPLICANT: Gilliam, Jacob T.
; TITLE OF INVENTION: Amino Polyl Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Unknown
US-09-352-159-31

Query Match          33.9%; Score 64.5; DB 4; Length 1196;
Best Local Similarity 54.2%; Pred. No. 6.1;
Matches 13; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 3 GPTLRQCL-AARAGGGGGGGIEG 25
   ||:| | | | | | | | | | |
Db 708 GPSIPPCADGAKAGGGGGGGSGG 731

RESULT 4
US-09-352-168-31
; Sequence 31, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyl Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
```

```
;
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Unknown
; US-09-352-168-31

Query Match          33.9%; Score 64.5; DB 4; Length 1196;
Best Local Similarity 54.2%; Pred. No. 6.1;
Matches 13; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 3 GPTLRQCL-AARAGGGGGGGIEG 25
   ||:| | | | | | | | | | |
Db 708 GPSIPPCADGAKAGGGGGGGSGG 731

RESULT 5
US-08-987-466-4
; Sequence 4, Application US/08987466
; Patent No. 5922595
; GENERAL INFORMATION:
; APPLICANT: Fisher, Douglas A.
; APPLICANT: Gooding, Doug
; APPLICANT: Streeter, Dave
; TITLE OF INVENTION: CYCLIC-GMP PHOSPHODIESTERASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,466
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0442 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 829179
; US-08-987-466-4

Query Match          33.7%; Score 64; DB 2; Length 584;
Best Local Similarity 61.9%; Pred. No. 3.5;
Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGIEGPTLRQC 31
   | | | | | | | | | | | |
Db 555 ALRAGGGGGGGGMAPRTGGC 575
```

US-09-100-564A-2
Sequence 2, Application US/09100664A
Patent No. 6057129
GENERAL INFORMATION:
APPLICANT: YOUNG, MICHAEL W.
APPLICANT: KLOSS, BRIAN
APPLICANT: BLAU, JUSTIN
APPLICANT: PRICE, JEFFREY
TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor

US-09-100-664A-3
; Sequence 3, Application US/09100664A
; Patent No. 6057129
; GENERAL INFORMATION:
; APPLICANT: YOUNG, MICHAEL W.
; APPLICANT: KLOSS, BRIAN
; APPLICANT: BLAU, JUSTIN
; APPLICANT: PRICE, JEFFREY
; TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,664A
; FILING DATE: 19-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-221
; TELECOMMUNICATION INFORMATION:

us-09-422-838c-28.ra1

Wed Oct 9 10:29:55 2002

```
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-100-664A-3

Query Match 32.6%; Score 62; DB 3; Length 440;
Best Local Similarity 55.0%; Pred. No. 4.4;
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLROCLAAARAGGGGGGI 23
Db 403 PERRPSIRMROGGGGGGV 422

RESULT 9
US-09-100-664A-4
Sequence 4, Application US/09100664A
Patent No. 6057129
GENERAL INFORMATION:
APPLICANT: YOUNG, MICHAEL W.
APPLICANT: KLOSS, BRIAN
APPLICANT: BLAU, JUSTIN
APPLICANT: PRICE, JEFFREY
TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,664A
FILING DATE: 19-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-221
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-100-664A-4

Query Match 32.6%; Score 62; DB 3; Length 440;
Best Local Similarity 55.0%; Pred. No. 4.4;
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLROCLAAARAGGGGGGI 23
Db 403 PERRPSIRMROGGGGGGV 422

TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-100-664A-3

Query Match 32.6%; Score 62; DB 3; Length 440;
Best Local Similarity 55.0%; Pred. No. 4.4;
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLROCLAAARAGGGGGGI 23
Db 403 PERRPSIRMROGGGGGGV 422

RESULT 10
US-08-764-640-13
Sequence 13, Application US/08764640
Patent No. 5869451
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 31.6%; Score 60; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAARA 14
Db 1 IEGPTLRQCLAAARA 14

RESULT 11
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
```


APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193

Query Match 31.6%; Score 60; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPLRLQCLAA 14
| | | | | | | | | |
Db 1 IEGPLRLQWLA 14

RESULT 12
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPLRLQCLAA 14
| | | | | | | | | |
Db 1 IEGPLRLQWLA 14

RESULT 13
US-08-973-225-193
Sequence 193, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

us-09-422-838c-28.ra1

Wed Oct 9 10:29:55 2002

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-08-973-225-193

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14
|||||

RESULT 14
US-09-244-298A-13
Sequence 13, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;

QY 1 IEGPTLRQCLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14
|||||

RESULT 15
US-09-244-298A-193
Sequence 193, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-193

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14
|||||

RESULT 16
US-09-516-704-13
Sequence 13, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.

;/ Cwirla, Steven E.
;/ Gates, Christian
;/ Schatz, Peter J.
;/ Balasubramanian, Palaniappan
;/ Wagstrom, Christopher R.
;/ Hendren, Richard W.
;/ Deprince, Randolph B.
;/ Podduturi, Surekha
;/ TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;/ RECEPTOR
;/ NUMBER OF SEQUENCES: 244
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Glaxo Wellcome
;/ STREET: Five Moore Drive, P.O. Box 13398
;/ CITY: Research Triangle Park
;/ STATE: NC
;/ COUNTRY: USA
;/ ZIP: 27709
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/516,704
;/ FILING DATE: 01-Mar-2000
;/ CLASSIFICATION: <Unknown>
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Hrubiec, Robert T.
;/ REGISTRATION NUMBER: 36,392
;/ REFERENCE/DOCKET NUMBER: PK3281
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 919-248-1000
;/ INFORMATION FOR SEQ ID NO: 13:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 14 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: <Unknown>
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 13:

US-09-516-704-13
Query Match 31.6%; Score 60; DB 4; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEGPTLRQCLAARA 14
| | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14
RESULT 17
US-09-516-704-193
; Sequence 193, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398

;/ CITY: Research Triangle Park
;/ STATE: NC
;/ COUNTRY: USA
;/ ZIP: 27709
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/516,704
;/ FILING DATE: 01-Mar-2000
;/ CLASSIFICATION: <Unknown>
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Hrubiec, Robert T.
;/ REGISTRATION NUMBER: 36,392
;/ REFERENCE/DOCKET NUMBER: PK3281
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 919-248-1000
;/ INFORMATION FOR SEQ ID NO: 193:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 14 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: <Unknown>
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-09-516-704-193
Query Match 31.6%; Score 60; DB 4; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEGPTLRQCLAARA 14
| | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 18
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996

us-09-422-838c-28.ra

Wed Oct 9 10:29:55 2002

```
;
; MOLECULE TYPE: peptide
; US-08-764-640-185
;
; Query Match 31.6%; Score 60; DB 2; Length 15;
; Best Local Similarity 92.9%; Pred. No. 0.27;
; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQCLAARA 14
; Db 2 IEGPTLRQWLAARA 15
;
; RESULT 20
; US-08-973-225-17
; ; Sequence 17, Application US/08973225A
; ; Patent No. 6083913
; ; GENERAL INFORMATION:
; ; APPLICANT: Dower, William J.
; ; Barrett, Ronald W.
; ; Cwirla, Steven E.
; ; Duffin, David J.
; ; Gates, Christian
; ; Haselden, Sherrill S.
; ; Mattheakis, Larry C.
; ; Schatz, Peter J.
; ; Wagstrom, Christopher R.
; ; Wrighton, Nicholas C.
; ; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; ; THROMBOPOIETIN RECEPTOR
; ; NUMBER OF SEQUENCES: 232
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Glaxo Wellcome
; ; STREET: Five Moore Drive, P.O. Box 13398
; ; CITY: Research Triangle Park
; ; STATE: NC
; ; COUNTRY: USA
; ; ZIP: 27709
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: PatentIn Release #1.0, Version #1.30
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/973,225A
; ; FILING DATE: 04-Dec-1997
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Hrubiec, Robert T.
; ; REGISTRATION NUMBER: 36,392
; ; REFERENCE/DOCKET NUMBER: PK3065USW
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: 919-248-1000
; ; INFORMATION FOR SEQ ID NO: 17:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 15 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS:
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: peptide
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
;
; US-08-973-225-17
;
; Query Match 31.6%; Score 60; DB 3; Length 15;
; Best Local Similarity 92.9%; Pred. No. 0.27;
; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQCLAARA 14
; Db 1 IEGPTLRQWLAARA 14
;
; RESULT 21
; US-08-973-225-185
; ; Sequence 185, Application US/08973225A
; ;
```

```
;
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-17
;
; Query Match 31.6%; Score 60; DB 2; Length 15;
; Best Local Similarity 92.9%; Pred. No. 0.27;
; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQCLAARA 14
; Db 1 IEGPTLRQWLAARA 14
;
; RESULT 19
; US-08-764-640-185
; ; Sequence 185, Application US/08764640
; ; Patent No. 5869451
; ; Patent No. 5869451 5837683
; ; GENERAL INFORMATION:
; ; APPLICANT: Dower, William J.
; ; Barrett, Ronald W.
; ; APPLICANT: Cwirla, Steven E.
; ; APPLICANT: Gates, Christian
; ; APPLICANT: Schatz, Peter J.
; ; APPLICANT: Balasubramanian, Palaniappan
; ; APPLICANT: Wagstrom, Christopher R.
; ; APPLICANT: Hendren, Richard W.
; ; APPLICANT: Deprence, Randolph B.
; ; APPLICANT: Podduturi, Surekha
; ; APPLICANT: Yin, Qun
; ; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; ; RECEPTOR
; ; NUMBER OF SEQUENCES: 244
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Glaxo Wellcome
; ; STREET: Five Moore Drive, P.O. Box 13398
; ; CITY: Research Triangle Park
; ; STATE: NC
; ; COUNTRY: USA
; ; ZIP: 27709
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: PatentIn Release #1.0, Version #1.30
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/764,640
; ; FILING DATE: 11-DEC-1996
; ; CLASSIFICATION: 514
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Hrubiec, Robert T.
; ; REGISTRATION NUMBER: 36,392
; ; REFERENCE/DOCKET NUMBER: PK3281
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: 919-248-1000
; ; INFORMATION FOR SEQ ID NO: 185:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 15 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS:
; ; TOPOLOGY: linear
; ;
```

Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-09-973-225-185
Query Match 31.6%; Score 60; DB 3; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 IEGPTLROCLAARA 14
||||||| |||||
Db 2 IEGPTLROWLAARA 15
RESULT 22
US-09-244-298A-17
Sequence 17, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17
Query Match 31.6%; Score 60; DB 3; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 IEGPTLROCLAARA 14
||||||| |||||
Db 1 IEGPTLROWLAARA 14
RESULT 23
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

TOPOLOGY: linear

RESULT 26

US-08-764-640-18
; Sequence 18, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depdince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
US-08-764-640-18

Query Match 31.6%; Score 60; DB 2; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEGPTLRQCLAARA 14
| | | | | | | | | |
Db 1 IEGPTLRQLAARA 14

RESULT 27
US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian

; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depdince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-764-640-194

Query Match 31.6%; Score 60; DB 2; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEGPTLRQCLAARA 14
| | | | | | | | | |
Db 2 IEGPTLRQLAARA 15

RESULT 28
US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depdince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: linear
MOLECULE TYPE: peptide
US-08-764-640-232

Query Match 31.6%; Score 60; DB 2; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 29
US-08-973-225-18
Sequence 18, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirila, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match 31.6%; Score 60; DB 3; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 30
US-08-973-225-194
Sequence 194, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirila, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 194:

Wed Oct 9 10:29:55 2002

US-08-973-225-194

Query Match 31.6%; Score 60; DB 3; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLARA 14
 |||||
Db 2 IEGPTLRQWLARA 15
 |||||

Search completed: October 9, 2002, 09:06:32
Job time : 6.98595 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8.09368 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838C-28
Perfect score: 190
Sequence: 1 IEPTLRQCLAAAGGGGGGIEPTLRQCLAAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	38.4	619	1 KSNCLC	laccase (EC 1.10.3
2	73	38.4	619	1 KSNCLT	laccase (EC 1.10.3
3	68	35.8	199	2 T48039	hypothetical prote
4	66	34.7	500	2 T20961	hypothetical prote
5	65.5	34.5	403	2 A53662	homeotic protein H
6	65	34.2	77	1 INSH	insulin precursor
7	65	34.2	105	1 IPBO	insulin precursor
8	65	34.2	201	2 JQ1094	hypothetical 20.2k
9	64.5	33.9	1733	1 B45344	probable nuclear a
10	64	33.7	331	2 T26807	hypothetical prote
11	64	33.7	333	2 T26808	hypothetical prote
12	64	33.7	777	2 S65543	3',5'-cyclic-nucle
13	63.5	33.4	434	2 T47772	hypothetical prote
14	63	33.2	488	2 G87033	probable ATP/Grp-b
15	63	33.2	518	2 S72938	hlx protein - Myc
16	63	33.2	806	2 T13690	hypothetical prote
17	63	33.2	1168	1 MWAXIC	myosin heavy chain
18	62.5	32.9	339	2 T06612	hypothetical prote
19	62	32.6	201	2 T49792	hypothetical prote
20	62	32.6	867	2 S57795	probable deoxyribo
21	62	32.6	889	2 T09055	protocadherin 68 -
22	61	32.1	495	2 D70505	probable Hflx - My
23	60	31.6	167	2 S71779	glycine-rich RNA-b
24	60	31.6	285	2 S69312	probable membrane
25	60	31.6	323	2 S20099	transforming prote
26	60	31.6	649	2 S58064	hdc protein - frui
27	60	31.6	1325	2 T13386	hypothetical prote
28	59.5	31.3	443	2 E96495	hypothetical prote
29	59.5	31.3	487	2 B39490	subtilisin-like pr

30 59.5 31.3 652 1 JC2191 subtilisin-like pr
31 59.5 31.3 962 2 JC5571 subtilisin-like pr
32 59.5 31.3 969 1 A39490 subtilisin-like pr
33 59.5 31.3 975 2 JC5570 subtilisin-like pr
34 59 31.1 102 2 H95992 hypothetical prote
35 59 31.1 165 2 S41773 glycine-rich RNA-b
36 59 31.1 165 2 S59529 RNA-binding glycin
37 59 31.1 250 2 H85067 hypothetical prote
38 59 31.1 298 2 C96690 unknown protein P2
39 59 31.1 346 1 S35500 heterogeneous ribo
40 59 31.1 367 2 JC6087 helix-loop-helix t
41 59 31.1 396 2 T49109 glycine-rich prote
42 59 31.1 517 2 B71260 hypothetical prote
43 59 31.1 543 2 F96624 hypothetical prote
44 59 31.1 593 1 KRH00 keratin 10, type I
45 59 31.1 1428 2 T13926 probable protein p

ALIGNMENTS

RESULT 1

KSNCLC
laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)
N:Alternate names: urishiol oxidase
C:Species: Neurospora crassa
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
C:Accession: A28523; A29762
R:Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
J. Biol. Chem. 263, 885-896, 1988
A:Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- a
A:Reference number: A28523; MUID:88087214
A:Accession: A28523
A:Molecule type: DNA
A:Residues: 1-619 <GER>
A:Cross-references: EMBL:M14554
R:Germann, U.A.; Lerch, K.
Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986
A:Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospor
A:Reference number: A29762; MUID:87067412
A:Accession: A29762
A:Molecule type: DNA
A:Residues: 379-619 <GE2>
A:Cross-references: GB:M14554; NID:gl68823; PIDN:AAA33590.1; PID:gl68824
C:Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquino
C:Genetics:
A:Introns: 86/3
C:Superfamily: laccase
C:Keywords: copper; glycoprotein; oxidoreductase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-49/Domain: propeptide #status predicted <PRO>
F:50-619/Product: laccase #status predicted <MAT>
F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
F:216-372/Domain: middle beta-barrel #status predicted <BB2>
F:439-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F:139,282,295,340,422,444/Binding site: carbonylhydride (Asn) (covalent) #status predict
F:144,480/Binding site: copper (His) (type 2) #status predicted
F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status p
F:147,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 38.4%; Score 73; DB 1; Length 619;
Best Local Similarity 60.0%; Pred. No. 1.2;
Matches 15; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7 RQCLARAGGGGGGIEPTLRQC 31
||| | ||||| |||
Db 39 RQDSQAFRYGGGGGCGNPTNRQC 63

RESULT 2

KSNCLT
laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)
N:Alternate names: urishiol oxidase

C:Species: Neurospora crassa
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
C:Accession: B28523
R:Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
J. Biol. Chem. 263, 885-896, 1988
A:Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
A:Reference number: A28523; MUID:88087214
A:Accession: B28523
A:Molecule type: DNA
A:Residues: 1-619 <GER>
A:Cross-references: EMBL:M19334; NID:g168827; PIDN:AAA33592.1; PID:g168828
C:Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
C:Genetics:
A:Introns: 86/3
C:Superfamily: laccase
C:Keywords: copper; glycoprotein; oxidoreductase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-49/Domain: propeptide #status predicted <PRO>
F:50-619/Product: laccase #status predicted <MAT>
F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
F:216-372/Domain: middle beta-barrel #status predicted <BB2>
F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F:139,282,295,340,422,444/Binding site: carboxylate (Asn) (covalent) #status predicted
F:144,480/Binding site: copper (His) (type 2) #status predicted
F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status pred
F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 38.4%; Score 73; DB 1; Length 619;
Best Local Similarity 60.0%; Pred. No. 1.2;
Matches 15; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7 RCLAAAGGGGGGGGIEGPTLRQC 31
|| | ||||| || |||
DB 39 RQDSQAERYGGGGGGCCNSPTNRQC 63

RESULT 3
T48099
hypothetical protein T20010.200 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48099
R:Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Rudd, S.; Lem
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24484
A:Accession: T48099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-199 <OBES>
A:Cross-references: EMBL:AL163816
A:Experimental source: cultivar Columbia; BAC clone T20010
C:Genetics:
A:Map position: 3
A:Introns: 163/2
A:Note: T20010.200

Query Match 35.8%; Score 68; DB 2; Length 199;
Best Local Similarity 34.8%; Pred. No. 1.5;
Matches 16; Conservative 3; Mismatches 11; Indels 16; Gaps 1;

QY 2 EGPTLRQC-----LAAGAGGGGGGIEGPTLRQC 31
|| | ||||| || |||
DB 7 EGRTRRCFPFASTCTLVAQTSLLCVDDGGGGGGVGGVDRGC 52

RESULT 4
T20961
hypothetical protein F1589.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20961
R:Percy, C.
submitted to the EMBL Data Library, August 1996

A:Reference number: Z19351
A:Accession: T20961
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-500 <WIL>
A:Cross-references: EMBL:Z78013; PIDN:CAB01420.1; GSPDB:GN00023; CESP:F15B9.5
A:Experimental source: clone F15B9
C:Genetics:
A:Gene: CESP:F15B9.5
A:Map position: 5
A:Introns: 46/3; 63/3; 125/2; 162/2; 283/3; 391/1; 446/1
Query Match 34.7%; Score 66; DB 2; Length 500;
Best Local Similarity 56.5%; Pred. No. 5.5;
Matches 13; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 GPTLROCLAAAGGGGGGGGIEG 25
| | | | | ||||| |
DB 429 GSNLGRFLSNRGGGGGGGGMG 451

RESULT 5
A53662
homeotic protein HB9 - human
C:Species: Homo sapiens (man)
C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 17-Oct-1997
C:Accession: A53662
R:Harrison, K.A.; Druey, K.M.; Dequchi, Y.; Tuscano, J.M.; Kehrl, J.H.
J. Biol. Chem. 269, 19568-19975, 1994
A:Title: A novel human homeobox gene distantly related to proboscipedia is expressed
A:Reference number: A53662; MUID:94327547
A:Accession: A53662
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-403 <HAR>
A:Cross-references: GB:U07663
A:Note: the nucleotide sequence and conceptual translation as given are self-consiste
C:Genetics:
A:Gene: GDB:HLXB9
A:Cross-references: GDB:I36411; OMIM:142994
A:Map position: 1q41-q42.1
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:244-300/Domain: homeobox homology <HOX>

Query Match 34.5%; Score 65.5; DB 2; Length 403;
Best Local Similarity 57.1%; Pred. No. 5.2;
Matches 16; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 10 LAARA---GGGGGGGIEGPTLRQCLAA 34
|| | ||||| | | |
DB 34 LAAAGTGGGGGGGASGGTSGSCSPA 61

RESULT 6
INSH
insulin precursor - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: S16430; S16431
R:Brown, H.; Sanger, F.; Kitai, R.
Biochem. J. 60, 556-565, 1955
A:Title: The structure of pig and sheep insulins.
A:Reference number: A90344
A:Accession: S16430
A:Molecule type: protein
A:Residues: 1-30, 57-77 <BRO>
R:Peterson, J.D.; Neirlich, S.; Oyer, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin
A:Reference number: A92111; MUID:72258016
A:Accession: S16431
A:Molecule type: protein

Query Match 33.4%; Score 63.5; DB 2; Length 434;
 Best Local Similarity 44.8%; Pred. No. 9;
 Matches 13; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

QY 13 RAGGGGGGG-----TEGPTLRQCL 32
 | | | | | | | | : | : | |
 Db 161 RVGGGGGGGDHVLVVEYVNGASLADCL 189

RESULT 14
 G87033
 probable ATP/GTP-binding protein [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: G87033
 R:Coile, S.F.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
 eam, M.A.; Ruthford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; MUID:21128732; PMID:11234002
 A:Accession: G87033
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-488 <STO>
 A:Cross-references: GB:AL450380; NID:gl3093026; PIDN:CAC31378.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: ML0997
 C:Superfamily: GTP-binding protein hflx; translation elongation factor Tu homology

Query Match 33.2%; Score 63; DB 2; Length 488;
 Best Local Similarity 45.8%; Pred. No. 11;
 Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 GPTLRQCLAAAGGGGGGIEGP 26
 | : : : | | | | | | : |
 Db 195 GESMSRQVGGRRGGGGGVLGRP 218

RESULT 15
 S72938
 hflx protein - Mycobacterium leprae
 A:Alternate names: B2235.C2.202 protein
 C:Species: Mycobacterium leprae
 C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
 C:Accession: S72938
 R:Smith, D.R.; Robison, K.
 submitted to the EMBL Data Library, November 1993
 A:Description: Mycobacterium leprae cosmid B2235.
 A:Reference number: S72938
 A:Accession: S72938
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-518 <SMI>
 A:Cross-references: EMBL:U00019; NID:g467079; PIDN:AAAL7274.1; PID:g467091
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: GTP-binding protein hflx; translation elongation factor Tu homology

Query Match 33.2%; Score 63; DB 2; Length 518;
 Best Local Similarity 45.8%; Pred. No. 12;
 Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 GPTLRQCLAAAGGGGGGIEGP 26
 | : : : | | | | | | : |
 Db 225 GESMSRQVGGRRGGGGGVLGRP 248

RESULT 16
 T13690
 hypothetical protein EG0003.2 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13690
 R:Murphy, L.; Harris, D.; Barrell, B.
 submitted to the EMBL Data Library, November 1998
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: Z17699
 A:Accession: T13690
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-806 <MUR>
 A:Cross-references: EMBL:AL031863; NID:el331652; PID:el355938; PIDN:CAA21318.1
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0025833
 A:Introns: 37/3; 448/3; 611/2; 690/3
 A:Note: EG:EG0003.2

Query Match 33.2%; Score 63; DB 2; Length 806;
 Best Local Similarity 54.5%; Pred. No. 17;
 Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEPTLRQCLAAARA 36
 | | | | | | | | : | : | |
 Db 100 GGGGGGGGPGGASITQAIQAQA 121

RESULT 17
 MWAXIC
 myosin heavy chain IC - Acanthamoeba castellanii
 N:Contains: myosin ATPase (EC 3.6.1.32)
 C:Species: Acanthamoeba castellanii
 C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
 C:Accession: A33891; C34448; A4146
 R:Jung, G.; Korn, E.D.; Hammer III, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
 A:Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like and non
 A:Reference number: A33891; MUID:88016163
 A:Accession: A33891
 A:Molecule type: DNA
 A:Residues: 1-1168 <JUN>
 A:Cross-references: GB:J02974; NID:gl55624; PIDN:AAA27707.1; PID:gl55625
 A:Note: This gene and protein are called MIB in this paper
 R:Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
 J. Biol. Chem. 264, 19340-19348, 1989
 A:Title: The localization and sequence of the phosphorylation sites of Acanthamoeba m
 A:Reference number: A34448; MUID:90037074
 A:Accession: C34448
 A:Molecule type: protein
 A:Residues: 308-314, 'X', 316-329 <BRZ>
 C:Comment: In this protein, the coiled-coil rod-like region found in many myosin heav
 he protein is globular and does not self-associate into filaments.
 C:Genetics:
 A:Gene: MIC
 A:Introns: 1/3; 37/3; 60/2; 100/2; 153/3; 179/3; 208/2; 242/3; 321/3; 371/3; 4
 C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 hom
 C:Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein
 F:10-653/Domain: myosin motor domain homology <MMOT>
 F:101-108/Region: nucleotide-binding motif A (P-loop)
 F:543-564/Region: actin binding #status predicted
 F:671-1168/Domain: carboxyl-terminal <CTD>
 F:675-883/Region: basic
 F:923-978/Region: alanine/glycine/proline-rich
 F:983-1030/Domain: SH3 homology <SH3>
 F:1034-1168/Region: alanine/glycine/proline-rich
 F:107/Binding site: ATP (Lys) #status predicted
 F:311/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 33.2%; Score 63; DB 1; Length 1168;
 Best Local Similarity 60.0%; Pred. No. 24;
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 8 QCLAAARAGGGGGGIEGPT 27
 | | | | | | | | : | : | |
 Db 920 QILGAKGGGGGGRGGGPS 939

RESULT 18
 6612
 hypothetical protein F16J13.120 - Arabidopsis thaliana
 Species: Arabidopsis thaliana (mouse-ear cress)
 Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 Accession: T06612
 Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, I.; Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, U. submitted to the Protein Sequence Database, April 1999
 Reference number: Z15789
 Accession: T06612
 Molecule type: DNA
 Residues: 1-339 <BEV>
 Cross-references: EMBL:AL049638; GSPDB:GN00062; ATSP:F16J13.120
 Experimental source: cultivar Columbia; BAC clone F16J13
 Genetics:
 Gene: ATSP:F16J13.120
 Map position: 4
 Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200
 Query Match 32.9% Score 62.5; DB 2; Length 339;
 Best Local Similarity 52.0% Pred. No. 9.3;
 Matches 13; Conservative 4; Mismatches 5; Indels 3; Gaps 1;
 y 15 GGGGGGGTGTGPTL---RQCLAARA 36
 ||||| 1 : : |||
 b 282 GGGGGGGGSPPMNGQQQAAAMA 306
 ||||| 1 : : |||
 RESULT 19
 49792
 hypothetical protein B9J10.290 [imported] - Neurospora crassa
 Species: Neurospora crassa
 Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 Accession: T49792
 Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, U. submitted to the Protein Sequence Database, May 2000
 Reference number: Z25022
 Accession: T49792
 Status: preliminary
 Molecule type: DNA
 Residues: 1-201 <SCH>
 Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.290
 Experimental source: BAC clone B9J10; strain OR74A
 Genetics:
 Gene: NCSP:B9J10.290
 Map position: 6
 Query Match 32.6% Score 62; DB 2; Length 201;
 Best Local Similarity 76.9% Pred. No. 6.7;
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 13 RAGGGGGGGGIEG 25
 | ||||| : |
 DB 74 RGGGGGGGGVNG 86
 ||||| : ||
 RESULT 20
 S57795
 Probable deoxyribodipyrimidine photo-lyase (EC 4.1.99.3) - Chlamydomonas reinhardtii
 N:Alternate names: DNA photolyase homolog; probable blue light photoreceptor
 C:Species: Chlamydomonas reinhardtii
 C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
 C:Accession: S57795; S66368
 R:Small, G.D.; Min, B.; Lefebvre, P.A.
 plant Mol. Biol. 28, 443-454, 1995
 A:Title: Characterization of a Chlamydomonas reinhardtii gene encoding a protein of the
 A:Reference number: S57795; MUID:95359403
 A:Accession: S57795
 A:Molecule type: DNA
 A:Residues: 1-867 <SMA>
 A:Cross-references: EMBL:L07561; NID:g945420; PID:945421

Db 205 GESMSRQAGGAGGGGGVGLRGP 228

RESULT 23

S71779

glycine-rich RNA-binding protein GRP1 - wheat

C:Species: Triticum aestivum (common wheat)

C>Date: 04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 23-Jul-1999

C:Accession: S71779

R:Gullitnan, M.J.; Niu, X.

Plant Mol. Biol. 30, 1301-1306, 1996

A:Title: cDNA encoding a wheat (Triticum aestivum cv. Chinese spring) glycine-rich RNA-binding protein

A:Reference number: S71779; MUID:96311016

A:Accession: S71779

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-167 <GUI>

A:Cross-references: EMBL:U32310; NID:g974604; PIDN:AAA75104.1; PID:g974605

C:Genetics:

A:Gene: GRP1

C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

F:7-74/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 31.6%; Score 60; DB 2; Length 167;

Best Local Similarity 55.6%; Pred. No. 9.3;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 TLROCLAAARAGGGGGGG 22

Db 78 TVNEAQSRSGGGGGGGG 95

RESULT 24

S69312

probable membrane protein YLR338w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L8300.13-a

C:Species: Saccharomyces cerevisiae

C>Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 05-Nov-1999

C:Accession: S69312

R:Du, Z.

submitted to the EMBL Data Library, January 1994

A:Description: The sequence of S. cerevisiae cosmid 8300.

A:Reference number: S69312

A:Accession: S69312

A:Molecule type: DNA

A:Residues: 1-285 <DUZ>

A:Cross-references: EMBL:U19028; NID:g609380; PID:g2340034; GSPDB:GN00012; MIPS:YLR338w

C:Genetics:

A:Gene: MIPS:YLR338w

A:Map position: 12R

C:Keywords: transmembrane protein

F:142-158/Domain: transmembrane #status predicted <TM1>

F:201-217/Domain: transmembrane #status predicted <TM2>

Query Match 31.6%; Score 60; DB 2; Length 285;

Best Local Similarity 57.9%; Pred. No. 15;

Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGGGIEGPTL 28

Db 236 LPPNAGGGGGGGGAGAPAI 254

RESULT 25

S20099

transforming protein jund - chicken

C:Species: Gallus gallus (chicken)

C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999

C:Accession: S20099

R:Hartl, M.; Hutchins, J.T.; Vogt, P.K.

Oncogene 6, 1623-1631, 1991

A:Title: The chicken jund gene and its product.

A:Reference number: S20099; MUID:92019832

A:Cross-references: FlyBase:FBgn0003575; FlyBase:FBgn0020381

A:Accession: S20099

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <HAR>

A:Cross-references: EMBL:X60063; NID:g62927; PIDN:CAA42665.1; PID:g62928

C:Superfamily: jun transforming protein; fos/jun DNA-binding domain homology

C:Keywords: DNA binding; nucleus; transcription regulation

F:237-277/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 31.6%; Score 60; DB 2; Length 323;

Best Local Similarity 72.2%; Pred. No. 16;

Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGGIEGPTL 28

Db 151 AAAAGGGGGGGGGGEL 168

RESULT 26

S58064

hdc protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 24-Sep-1998

C:Accession: S58064

R:Weaver, T.A.; White, R.A.

submitted to the EMBL Data Library, July 1995

A:Description: hdc, an imaginal specific gene required for adult morphogenesis in Dro

A:Reference number: S58064

A:Accession: S58064

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-649 <WEA>

A:Cross-references: EMBL:Z50097; NID:g902623; PID:g902624

C:Genetics:

A:Gene: FlyBase:hdc

A:Cross-references: FlyBase:FBgn0010113

Query Match 31.6%; Score 60; DB 2; Length 649;

Best Local Similarity 76.9%; Pred. No. 30;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GGGGGGGGGGIEGPT 27

Db 220 GGGGGGGGGVNGNT 232

RESULT 27

TL3386

hypothetical protein l15C2.3 - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: TL3386; A39612

R:Salles, C.; Valenti, P.; Darlamitsou, A.; Henderson, N.; Campbell, L.; Glover, D.

submitted to the EMBL Data Library, May 1999

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: Z17665

A:Accession: TL3386

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1325 <CAT>

A:Cross-references: EMBL:AL031581; NID:el320978; PID:el320992; PIDN:CAA20886.1

R:Voelker, R.A.; Gibson, W.; Graves, J.P.; Sterling, J.F.; Eisenberg, M.T.

Mol. Cell. Biol. 11, 894-905, 1991

A:Title: The Drosophila suppressor of sable gene encodes a polypeptide with regions s

A:Reference number: A39612; MUID:91117256

A:Accession: A39612

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-719,723-1325 <VOE>

A:Cross-references: GB:M57889; NID:gl58516; PID:gl58517

C:Genetics:

A:Gene: FlyBase:su(s)

us-09-422-838c-28.rpr

Wed Oct 9 10:29:56 2002

Query Match 31.3%; Score 59.5; DB 2; Length 487;
 Best Local Similarity 60.0%; Pred. No. 27;
 Matches 15; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 11 AARAGGGGGGGIEGPTLRQCLAAAR 35
 ||||| ||||| ||||| ||||| |||||
 Db 24 AAGAGGAGGAGGAGGAGGFRP-LAPR 47

RESULT 30

JC2191
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form C -
 N:Alternate names: kexin-like protease isoform
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 20-Apr-2000
 C:Accession: JC2191
 R:Tsuji, A.; Higashine, K.; Hine, C.; Mori, K.; Tamai, Y.; Nagamune, H.; Matsuda, Y.
 Biochem. Biophys. Res. Commun. 200, 943-950, 1994
 A:Title: Identification of novel cDNAs encoding human kexin-like protease, PACE4 isof
 A:Reference number: JC2191; MUID:94235049
 A:Accession: JC2191
 A:Molecule type: mRNA
 A:Residues: 1-652 <TSU>
 C:Comment: This protein consists of a signal peptide, a propeptide, a subtilisin-like
 C:Comment: This protein cleaves precursor proteins at dibasic amino acid residues.
 C:Genetics:
 A:Gene: GDB:PACE4
 A:Cross-references: GDB:131390; OMIM:167405
 A:Map position: 15q26-15q26
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C:Keywords: alternative splicing; hydrolase; serine proteinase
 F:196-434/Domain: subtilisin homology <SPT>
 F:205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 31.3%; Score 59.5; DB 1; Length 652;
 Best Local Similarity 60.0%; Pred. No. 34;
 Matches 15; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 11 AARAGGGGGGGIEGPTLRQCLAAAR 35
 ||||| ||||| ||||| ||||| |||||
 Db 24 AAGAGGAGGAGGAGGAGGFRP-LAPR 47

Search completed: October 9, 2002, 09:05:07
 Job time : 9.09368 secs

A:Map position: X
 A:Introns: 92/1; 170/3; 603/2; 645/1
 A:Note: EG-115C2.3

Query Match 31.6%; Score 60; DB 2; Length 1325;
 Best Local Similarity 58.8%; Pred. No. 56;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 15 GGGGGGGGGIEGPTLRQ 30
 ||||| ||||| ||||| ||||| |||||
 Db 1162 GGGGGGGGGVLPNLSQ 1177

RESULT 28

E96495
 hypothetical protein F8D11.2 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
 C:Accession: E96495
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E96495
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-443 <STO>
 A:Cross-references: GB:AE005173; NID:gi10092398; PIDN:AAG12804.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F8D11.2
 A:Map position: 1
 C:Superfamily: barley pathogen resistance protein M10

Query Match 31.3%; Score 59.5; DB 2; Length 443;
 Best Local Similarity 92.3%; Pred. No. 24;
 Matches 12; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 14 AGGGGGGGGIEGP 26
 ||||| ||||| ||||| ||||| |||||
 Db 2 AGGGGGGGG-EGP 13

RESULT 29

B39490
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form B - hum
 N:Alternate names: subtilisin homolog precursor, short splice form
 C:Species: Homo sapiens (man)
 C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 31-Mar-2000
 C:Accession: B39490
 R:Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landsberg, K.E.; Saltman, D.; Barr, P.J.
 DNA Cell Biol. 10, 757-769, 1991
 A:Title: Identification of a second human subtilisin-like protease gene in the fes/fps
 A:Reference number: A39490; MUID:92075167
 A:Accession: B39490
 A:Molecule type: mRNA
 A:Residues: 1-487 <KIE>
 A:Note: The lack of a domain necessary for correct folding and activity of other serine
 C:Genetics:
 A:Gene: GDB:PACE4
 A:Cross-references: GDB:131390; OMIM:167405
 A:Map position: 15q26-15q26
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C:Keywords: alternative splicing; hydrolase; serine proteinase
 F:196-434/Domain: subtilisin homology <SPT>
 F:205,246,420/Active site: Asp, His, Ser #status predicted

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-28

Perfect score: 190

Sequence: 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	38.4	619	1	LAC1_NEUCR
2	73	38.4	619	1	LAC2_NEUCR
3	65	34.2	105	1	INS_BOVIN
4	65	34.2	201	1	YR21_TRSVR
5	65	34.2	401	1	H99_HUMAN
6	65	34.2	1263	1	SVV2_MOUSE
7	64.5	33.9	1733	1	VNUA_PPVKA
8	64	33.7	105	1	INS_SHEEP
9	64	33.7	584	1	CNAL_DROME
10	63	33.2	394	1	FXD3_CHICK
11	63	33.2	1168	1	MYSC_ACACA
12	62	32.6	440	1	DCO_DROME
13	61.5	32.4	4499	1	DYHA_CHLRE
14	61	32.1	1178	1	PHYB_SORBI
15	60	31.6	323	1	JUND_CHICK
16	60	31.6	348	1	SXL_CERCA
17	60	31.6	1322	1	SUS_DROME
18	59.5	31.3	391	1	SOX1_MOUSE
19	59.5	31.3	969	1	PAC4_HUMAN
20	59	31.1	367	1	BET3_MESAU
21	59	31.1	497	1	FXD2_HUMAN
22	59	31.1	517	1	I967_TREPA
23	59	31.1	593	1	KLCU_HUMAN
24	59	31.1	753	1	ZIN_HUMAN
25	59	31.1	757	1	ECR_LUCCU
26	58.5	30.8	168	1	SSB_MYCLE
27	58	30.5	445	1	H3R_HUMAN
28	58	30.5	476	1	EVX2_HUMAN
29	58	30.5	485	1	ONC2_HUMAN
30	58	30.5	495	1	BRN1_MOUSE
31	58	30.5	497	1	BRN1_RAT
32	58	30.5	500	1	BRN1_HUMAN
33	58	30.5	569	1	KLCU_MOUSE

34	58	30.5	644	1	XYND_CELFI
35	58	30.5	688	1	BOMD_MOUSE
36	58	30.5	796	1	KF3C_RAT
37	58	30.5	1171	1	PHYB_ORISA
38	58	30.5	3703	1	ABFI_HUMAN
39	57.5	30.3	368	1	ST19_HUMAN
40	57	30.0	266	1	CANS_RAT
41	57	30.0	269	1	CANS_MOUSE
42	57	30.0	301	1	CCO2_CAEEL
43	57	30.0	369	1	TMAF_AVIS4
44	57	30.0	375	1	PER_DROSC
45	57	30.0	440	1	FXGA_CHICK

ALIGNMENTS

RESULT 1
LAC1_NEUCR
ID LAC1_NEUCR STANDARD; PRT: 619 AA.
AC P06811;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
DE (Urishiol oxidase) (Laccase allele OR).
GN LACC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88087214; PubMed=2961749;
RA Hermann U.A., Mueller G., Hunziker P.E., Lerch K.;
RT "Characterization of two allelic forms of Neurospora crassa laccase.
RT Amino- and carboxyl-terminal processing of a precursor.";
RL J. Biol. Chem. 263:885-896(1988).
RN [2]
RP SEQUENCE OF 379-619 FROM N.A.
RX MEDLINE=87067412; PubMed=2947240;
RA Hermann U.A., Lerch K.;
RT "Isolation and partial nucleotide sequence of the laccase gene from
RT Neurospora crassa: amino acid sequence homology of the protein to
RT human ceruloplasmin".
RL Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).
CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLY).
CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.
CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M14554; AAA33590.1;
CC EMBL: M18333; AAA33591.1;
CC PIR: A28523; KSNCL0.
CC PIR: A29762; A29762.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR002355; MultiCu_oxidse2.
CC Pfam: PF00394; Cu-oxidase; 3
CC PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
DR

PROSTRE; PS00080; MULTICOPPER_OXIDASE2: 1;
Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
Glycoprotein; Repeat.

[illegible]

LAC2_NEUCR			
ID	LAC2_NEUCR	STANDARD;	PRT; 619 AA.
DT	PI0574;		
DT	01-JUL-1989	(Rel. 11, Created)	
DT	01-FEB-1996	(Rel. 33, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Laccase precursor	(EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)	
DE	Laccase precursor	(EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)	
DE	Laccase precursor	(EC 1.10.3.2) (Laccase allelic TS).	

LACC.
 Neurospora crassa.
 Eukaryote: Fungi: Ascomycota: Pezizomycotina: Sordariomycetes;
 OC Sordariales: Sordariaceae: Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A. MEDLINE=85087214; PubMed=2961749;
 RX German U.A., Mueller G., Hunziker P.E., Lerch K.;
 RA "Characterization of two allelic forms of Neurospora crassa laccase.
 FT Amino- and carboxyl-terminal processing of a precursor.";
 RL J. Biol. Chem. 263:885-896(1988).
 CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2
 CC H(2)O.
 CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC with the EMBL, our station

Wed Oct 9 10:29:57 2002

```

FT DOMAIN 144 148 POLY-GLY.
SQ SEQUENCE 201 AA; 20194 MW; 9038506E18D7B450 CRC64;

Query Match 34.2%; Score 65; DB 1; Length 201;
Best Local Similarity 61.5%; Pred. No. 1.7;
Matches 16; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 13 RAGGGGGGGGIE-----GPTLRQCLAA 34
   | | | | | | | | | | | | | | | | | |
Db 13 RAGGGGGGGGKEVFKAGRTLLKVLKA 38

RESULT 5
ID HB9_HUMAN STANDARD; PRT; 401 AA.
AC P50219;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Homeobox protein HB9.
GN HLB9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94327347; PubMed=7914194;
RA "A novel human homeobox gene distantly related to proboscipedia is
RT expressed in lymphoid and pancreatic tissues.";
RL J. Biol. Chem. 269:19968-19975(1994).
CC -!- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
CC -!- SIMILARITY: TO DROSOPHILA HOMEOBOX PROTEIN PROBOSCIPEDIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U07664; AAB60647.1; -.
CC EMBL; U07663; AAB60647.1; JOINED.
CC HSP; P14653; I872.
CC TRANSFAC; T03420; -.
CC MIM; 142994; -.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 39 48 POLY-GLY.
FT DOMAIN 97 111 POLY-GLY.
FT DOMAIN 120 135 POLY-ALA.
FT DOMAIN 169 177 POLY-ALA.
FT DNA_BIND 242 301 HOMEBOX.
FT DOMAIN 316 325 POLY-GLY.
SQ SEQUENCE 401 AA; 40932 MW; 0006AEAD71D594FE CRC64;

Query Match 34.2%; Score 65; DB 1; Length 401;
Best Local Similarity 54.2%; Pred. No. 3;
Matches 13; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGTGGTPTLRQCLAA 34
   | | | | | | | | | | | | | | | | | |
Db 37 ASGTGGGGGGGNSGTSCTSCSPA 60

```

```

RESULT 6
SV2_MOUSE STANDARD; PRT; 1263 AA.
ID SV2_MOUSE
AC Q92109; Q9QUN2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Valyl-tRNA synthetase 2 (EC 6.1.1.9) (Valine--tRNA ligase 2) (VALRS
DE 2).
GN VARS2 OR G7A OR BAT6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Rowen L., Madan A., Qin S., Shaffer T., Ratcliffe A., Abbasi N.,
RA Dickhoff R., James R., Loretz C., Lasky S., Hood L.;
RA "Sequence of the mouse major histocompatibility locus class III
RT region.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, AND C57BL/RIJ; TISSUE=Brain;
RX MEDLINE=99216447; PubMed=10199925;
RA "The sequence and organization of the mouse valyl-tRNA synthetase gene
RT G7a/Bat6 located in the MHC class III region.";
RL Immunogenetics 49:468-470(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC + L-valyl-tRNA(Val).
CC -!- ENZYME REGULATION: CAN BE REGULATED BY PROTEIN KINASE C-DEPENDENT
CC PHOSPHORYLATION (BY SIMILARITY).
CC -!- SUBUNIT: FORMS HIGH-MOLECULAR-MASS AGGREGATES WITH ELONGATION
CC FACTOR 1 (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -!- SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION
CC FACTOR 1-GAMMA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF109905; AAC84151.1; -.
CC EMBL; AF109906; AAC84172.1; -.
CC EMBL; AF087680; AAD26532.1; -.
CC EMBL; AF087141; AAD26531.1; -.
CC HSP; P96142; IGAX.
CC MGD; P9190675; Vars2.
CC InterPro; IPR004046; GST-C.
CC InterPro; IPR004045; GST-N.
CC InterPro; IPR002300; tRNA-synt_la.
CC InterPro; IPR001412; tRNA-synt_I.
CC InterPro; IPR002303; tRNA-synt_val.
CC Pfam; PF00043; GST_C; 1.
CC Pfam; PF02798; GST_N; 1.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00986; TRNASYNTHAL.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 1 2200 EF-1-GAMMA LIKE.
FT SITE 343 353 "HIGH" REGION.
FT SITE 861 863 "KMSKS" REGION.
FT BINDING 864 864 ATP (BY SIMILARITY).
FT CONFLICT 959 959 A -> R (IN REF. 2).

```

```

OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94280618; PubMed=8011164;
RA Ohlsen S.M., Lucenbeel K.A., Wong E.A.;
RT "Characterization of the linked ovine insulin and insulin-like growth factor-II genes.";
RL DNA Cell Biol. 13:377-388(1994).
RN [2]
RP SEQUENCE OF 25-54 AND 85-105.
RA Brown H., Sanger F., Kitai R.;
RT "The structure of pig and sheep insulins.";
RL Biochem. J. 60:556-565(1955).
RN [3]
RP SEQUENCE OF 57-82.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrllich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin C-peptides by a semi-micro Edman degradation procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC [-] FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC [-] SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC [-] SUBCELLULAR LOCATION: Secreted.
CC [-] SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to licens@isb-sib.ch)
CC EMBL; U00659; AAB60625.1; -.
DR PIR; S16430; INSH.
DR HSPP; P01315; 9INS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL
FT CHAIN
FT PROPEP
FT CHAIN
FT DISULFID
FT DISULFID
FT DISULFID
SQ SEQUENCE 105 AA; 11235 MW; 8E27C7FB9992BC7A CRC64;
Query Match 33.7%; Score 64; DB 1; Length 105;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches 16; Conservative 3; Mismatches 11; Indels 8; Gaps
QY 1 IGSPVLRQLAARAGGGGSGIEGP-----TLRQCILA 33
DB 58 VEGP---QVGALETAGPGAGGLEGGPKRGIVQQCA 92
RESULT 9
ID _CNAL_DROME
AC P12252; STANDARD; PRT; 584 AA.
DT 01-OCT-1989 (Rel. 12, Created)
DD 01-OCT-1996 (Rel. 34, Last sequence update)
```


Wed Oct 9 10:29:57 2002

FT NP_BIND 15 23 ATP (BY SIMILARITY).
 FT ACT_SITE 38 38 ATP (BY SIMILARITY).
 FT DOMAIN 128 128 BY SIMILARITY.
 FT DOMAIN 319 332 POLY-ALA.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 351 POLY-GLY.
 FT DOMAIN 414 426 POLY-GLY.
 FT DOMAIN 430 437 POLY-GLY.
 FT MUTAGEN 47 47 PERIOD.
 FT MUTAGEN 80 80 M->I: IN DBTL; LENGTHENS THE BEHAVIORAL PERIOD.
 FT MUTAGEN 80 80 M->I: IN DBTL; LENGTHENS THE BEHAVIORAL PERIOD.
 SQ SEQUENCE 440 AA; 48073 MW; B875891D5747391D CRC64;
 Query Match 32.6%; Score 62; DB 1; Length 440;
 Best Local Similarity 55.0%; Pred. No. 6.6;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 4 PTLRQCLAAARAGGGGGGGI 23
 DB 403 PRRPSIRMQGGGGGGV 422
 RESULT 13
 ID DYHA_CHLRE STANDARD; PRT; 4499 AA.
 AC Q39610;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dynein alpha chain, flagellar outer arm (DHC alpha).
 GN ODALL OR ODA-11.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A., AND REVISIONS.
 RC STRAIN=21GR;
 RX MEDLINE=97329535; PubMed=9186009;
 RA Mitchell D.R., Brown K.S.;
 RT "Sequence analysis of the Chlamydomonas reinhardtii flagellar alpha dynein gene."
 RL Cell Motil. Cytoskeleton 37:120-126(1997).
 RN [2]
 RP SEQUENCE OF 1142-4499 FROM N.A.
 RC STRAIN=21GR;
 RX MEDLINE=94274778; PubMed=8006077;
 RA Mitchell D.R., Brown K.S.;
 RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy chain genes."
 RL J. Cell Sci. 107:635-644(1994).
 CC -1- FUNCTION: PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
 CC FLAGELLA.
 CC DYNEIN HAS ATPASE ACTIVITY.
 CC -1- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L26049; AAA57316.2;
 CC InterPro; IPR003593; AAA.
 CC InterPro; IPR001298; Filamin.
 CC InterPro; IPR002909; IPT_TIG.
 CC InterPro; IPR001798; Kelch.
 CC InterPro; IPR001736; PLD.

DR Pfam; PF00630; Filamin; 1.
 DR Pfam; PF01344; Kelch; 3.
 DR SMART; SM00382; AAA; 3.
 DR SMART; SM00429; IPT; 1.
 DR PROSITE; PS0194; FILAMIN_REPEAT; 1.
 KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
 KW Coiled coil.
 FT REPEAT 425 534 FILAMIN.
 FT DOMAIN 1261 1334 COILED COIL (POTENTIAL).
 FT DOMAIN 1382 1450 COILED COIL (POTENTIAL).
 FT DOMAIN 1836 1864 MICROTUBULE-BINDING (POTENTIAL).
 FT DOMAIN 2655 2698 COILED COIL (POTENTIAL).
 FT DOMAIN 3003 3023 COILED COIL (POTENTIAL).
 FT DOMAIN 3170 3262 COILED COIL (POTENTIAL).
 FT DOMAIN 3486 3515 COILED COIL (POTENTIAL).
 FT NP_BIND 1716 1723 ATP (POTENTIAL).
 FT NP_BIND 2019 2026 ATP (POTENTIAL).
 FT NP_BIND 2369 2376 ATP (POTENTIAL).
 FT NP_BIND 2717 2754 ATP (POTENTIAL).
 SQ SEQUENCE 4499 AA; 503606 MW; 319AC7FD30F1591A CRC64;
 Query Match 32.4%; Score 61.5; DB 1; Length 4499;
 Best Local Similarity 48.5%; Pred. No. 54;
 Matches 16; Conservative 3; Mismatches 11; Indels 3; Gaps 1;
 QY 3 GPTLRQCLAAARAGGGGGGGI 32
 DB 4194 GETLFTVVEVAGGGGGGGGGG 4226
 RESULT 14
 ID PHVB_SORBI STANDARD; PRT; 1178 AA.
 AC P93527;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Phytochrome B.
 GN PHVB OR MA3.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. 58M;
 RX MEDLINE=20188796; PubMed=10723737;
 RA Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;
 RT "The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms."
 RL Mol. Biol. Evol. 17:362-373(2000).
 RN [2]
 RP SEQUENCE OF 208-1178 FROM N.A.
 RC STRAIN=CV. 58M;
 RX MEDLINE=97198556; PubMed=9046599;
 RA Childs K.L., Miller F.R., Cordonnier-Pratt M.-M., Pratt L.H., Morgan P.W., Mullet J.E.;
 RT "The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a phytochrome B."
 RL Plant Physiol. 113:611-619(1997).
 CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE PR FORM THAT ABSORBS MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN, PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION (BY SIMILARITY).

CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPEPTIDE CHROMOPHORE.
 CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF182394; AAB41398.2; .
 DR InterPro; IPR003018; GAF.
 DR InterPro; IPR003594; HATPase_c.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR InterPro; IPR003661; HIS_KIN.
 DR InterPro; IPR000014; PAS.
 DR InterPro; IPR001294; Phytochrome.
 DR Pfam; PF01590; GAF; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR Pfam; PF00989; PAS; 2.
 DR Pfam; PF00360; phytochrome; 1.
 DR Pfam; PF00512; signal; 1.
 DR PRINTS; PR01033; PHYTOCHROME.
 DR SMART; SM00065; GAF; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR SMART; SM00388; HSKA; 1.
 DR SMART; SM00091; PAS; 2.
 DR PROSITE; PS00109; HIS_KIN; 1.
 DR PROSITE; PS01112; PAS; 2.
 DR PROSITE; PS00245; PHYTOCHROME_1; 1.
 DR PROSITE; PS50046; PHYTOCHROME_2; 1.
 CC Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
 KW Repeat; Multigene family.
 FT DOMAIN 668 739 PAS 1.
 FT DOMAIN 802 873 PAS 2.
 FT DOMAIN 950 1170 HISTIDINE KINASE.
 FT DOMAIN 23 31 POLY-HIS.
 FT DOMAIN 43 54 POLY-GLY.
 FT BINDING 372 372 CHROMOPHORE (BY SIMILARITY).
 SQ SEQUENCE 1178 AA; 129136 MW; C406DF221197B93F CRC64;
 Query Match 32.1%; Score 61; DB 1; Length 1178;
 Best Local Similarity 75.0%; Pred. No. 19;
 Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 12 ARAGGGGGGGGIEGPT 27
 Db :||||||||| |
 40 SRAGGGGGGGGGGGGT 55
 [1]
 RESULT 15
 JUND_CHICK
 ID JUND_CHICK STANDARD; PRT; 323 AA.
 AC P27921;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Transcription factor jund-D.
 GN JUND.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92019832; PubMed=1923529;
 RA Hartl M., Hutchins J.T., Vogt P.K.;
 RT "The chicken jund gene and its product.";

RL Oncogene 6:1623-1631(1991).
 CC -!- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. JUN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X60063; CAA42665.1; .
 DR PIR; S20099; S20099.
 DR HSP; P05412; 1FOS.
 DR TRANSFAC; T02196; .
 DR InterPro; IPR002112; Leuzip-Jun.
 DR InterPro; IPR001871; bZIP.
 DR Pfam; PF00170; bZIP; 1.
 DR PRINTS; PR00043; LEUZIPPRJUN.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
 FT DOMAIN 59 67 POLY-ALA.
 FT DOMAIN 155 166 POLY-GLY.
 FT DNA_BIND 242 266 BASIC MOTIF.
 FT DOMAIN 270 298 LEUCINE-ZIPPER.
 SQ SEQUENCE 323 AA; 33205 MW; A7F6D21A97DB8676 CRC64;
 Query Match 31.6%; Score 60; DB 1; Length 323;
 Best Local Similarity 72.2%; Pred. No. 8.2;
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 11 AARAGGGGGGGIEGPTL 28
 Db :||||||||| |
 151 AAAAGGGGGGGGGGSEL 168
 [1]
 RESULT 16
 SXL_CERCA
 ID SXL_CERCA STANDARD; PRT; 348 AA.
 AC O61374;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sex-lethal protein homolog (CCSXL).
 GN SXL.
 OS Ceratitis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritoidea; Tephritidae; Ceratitis.
 CC NCBI_TaxID=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BENAKIO;
 RX MEDLINE=98171464; PubMed=9502730;
 RA Saccone G., Peluso I., Artalejo D., Giordano E., Bopp D., Polito L.C.;
 RT "The Ceratitis capitata homologue of the Drosophila sex-determining
 RT gene Sex-lethal is structurally conserved, but not sex-specifically
 RT regulated.";
 RL Development 125:1495-1500(1998).
 CC -!- FUNCTION: UNKNOWN; APPARENTLY NOT INVOLVED IN SOMATIC SEX
 CC DETERMINATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS; ADULT-SPECIFIC ISOFORMS
 CC A1, A2, A3, A4, AND EMBRYO-SPECIFIC ISOFORMS E1, E2 AND E3 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS OF BOTH SEXES. ALSO
 CC EXPRESSED IN THE PROGENITOR CELLS OF THE GERMLINE.
 CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF026145; AAC38968.1; -
CC HSP; P19339; 1SXL.
CC InterPro: IPR000504; RRM.
CC Pfam: PF00076; rrm; 2.
CC PRINTS: PR00961; HUDSLRNA.
CC SMART; SM00360; RRM; 2.
CC PROSITE; PS00102; RRM; 2.
CC PROSITE; PS00030; RRM_RNP_1; 1.
CC RNA-binding; Repeat; Nuclear GLY/ASN-RICH DOMAIN.
CC FT DOMAIN 1 27
CC FT DOMAIN 110 188
CC FT DOMAIN 196 276
CC FT DOMAIN 68 75
CC FT DOMAIN 95 99
CC FT DOMAIN 293 311
CC FT DOMAIN 312 316
CC FT VARSPLIC 37 44
CC SEQUENCE 348 AA; 37188 MW; CABA3DA5C2G8874A CRC64;
Query Match 31.6%; Score 60; DB 1; Length 348;
Best Local Similarity 83.3%; Pred. No. 8.7; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 1; Mismatches 0; Gaps 0;
QY 15 GGGGGGGGIEGP 26
DB 301 GGGGGGGGNGGP 312
RESULT 17
SUS_DROME
ID SUS_DROME STANDARD; PRT; 1322 AA.
AC P22293;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Suppressor of sable protein.
GN SU(S).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=91117256; PubMed=1703632;
RA Voelker R.A., Gibson W., Graves J.P., Sterling J.F., Eisenberg M.T.;
RT "The Drosophila suppressor of sable gene encodes a polypeptide with
RL regions similar to those of RNA-binding proteins.",
RL Mol. Cell. Biol. 11:894-905(1991).
RN [2]
RP SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE=91169252; PubMed=1963868;
RA Voelker R.A., Graves J.P., Gibson W., Eisenberg M.T.;
RT "Mobile element insertions causing mutations in the Drosophila
RT suppressor of sable locus occur in DNase I hypersensitive subregions
RT of 5'-transcribed nontranslated sequences.",
RL Genetics 126:1071-1082(1990).
CC -1- FUNCTION: AFFECTS THE TRANSCRIPT LEVELS OF THOSE ALLELES THAT IT
CC SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
CC -1- SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: AT ALL STAGES.
CC -1- SIMILARITY: HAS REGIONS SIMILAR TO THOSE OF RNA-BINDING PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC -----
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M57889; AAA28920.1; -
CC EMBL; X59364; CAA42010.1; -
CC PIR; A39612; A39612.
CC FlyBase; FBgn0003575; su(s).
CC InterPro: IPR000571; Zf-CCCH; 2.
CC Pfam: PF00642; zf-CCCH; 2.
CC RNA-binding; Nuclear protein.
CC FT DOMAIN 138 327
CC FT DOMAIN 446 474
CC FT DOMAIN 1087 1162
CC SEQUENCE 1322 AA; 143555 MW; D5F534EB5702EA08 CRC64;
Query Match 31.6%; Score 60; DB 1; Length 1322;
Best Local Similarity 68.8%; Pred. No. 27;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 15 GGGGGGGGIEGPTLRQ 30
DB 1159 GGGGGGGGVLPNLSQ 1174
RESULT 18
SOX1_MOUSE
ID SOX1_MOUSE STANDARD; PRT; 391 AA.
AC P53783;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SOX-1 protein.
GN SOX1 OR SOX-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=96189340; PubMed=8625802;
RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,
RA Lovell-Badge R.;
RT "A comparison of the properties of Sox-3 with Sry and two related
RT genes, Sox-1 and Sox-2.",
RL Development 122:509-520(1996).
CC -1- SUBCELLULAR LOCATION: Nuclear (probable).
CC -1- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
CC SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X94126; CAA63846.1; -
CC HSP; Q05066; 1HRY.
CC MGD; MGI:98357; Sox1.
CC InterPro: IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 1.
CC SMART; SM00398; HMG; 1.
CC DNA-binding; Nuclear protein.
CC FT DOMAIN 30 43
CC FT DNA_BIND 51 119
CC FT DOMAIN 145 150
CC POLY-GLY.
CC HMG BOX.
CC POLY-GLY.

FT DOMAIN 197 204 POLY-ALA.
 FT DOMAIN 280 288 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 357 364 POLY-ALA.
 SQ SEQUENCE 391 AA; 392337 MW; 9F81ED667F947C05 CRC64;
 Query Match 31.3%; Score 59.5; DB 1; Length 391;
 Best Local Similarity 54.5%; Pred. No. 11;
 Matches 12; Conservative 1; Mismatches 4; Indels 5; Gaps 1;
 QY 1 IEPTTLRQCLAAARAGGGGGG 22
 Db 22 LSGPA-----GARGGGGGGG 38
 RESULT 19
 PAC4_HUMAN
 ID PAC4_HUMAN STANDARD: PRT: 969 AA.
 AC Q29122; Q15099; Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
 AC Q9UEG7; Q9Y4G9; Q9Y4H0; Q9Y4H1;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like propeptide
 DE convertase 4) (SPC4).
 GN PACE4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).
 RC TISSUE=Hepatosoma, and Kidney;
 RX MEDLINE=92075167; PubMed=1741956;
 RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Sallman D.,
 RA Barr P.J.;
 RT "Identification of a second human subtilisin-like protease gene in
 RT the fes/fps region of chromosome 15.";
 RL DNA Cell Biol. 10:757-769(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).
 RC TISSUE=Placenta;
 RX MEDLINE=94235049; PubMed=8179631;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms.";
 RL Biochem. Biophys. Res. Commun. 200:943-950(1994).
 RN [3]
 RP ERRATUM.
 RX MEDLINE=95071480; PubMed=7980617;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms.";
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM PACE4A-II).
 RC TISSUE=Placenta;
 RX Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
 RA Matsuda Y.;
 RT "Identification of a novel PACE4 isoform, PACE4E.";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).
 RC TISSUE=Cerebellum;
 RX MEDLINE=97335942; PubMed=9192737;
 RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
 RA Akamatsu T., Nagamune H., Matsuda Y.;
 RT "A novel human PACE4 isoform, PACE4E is an active processing protease
 RT containing a hydrophobic cluster at the carboxy terminus.";
 RL J. Biochem. 121:941-948(1997).

RN [6]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; A-II; CS; D; E-I; E-II).
 RX MEDLINE=98021085; PubMed=9378725;
 RA Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
 RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
 RT "Genomic organization and alternative splicing of human PACE4 (SPC4),
 RT kexin-like processing endoprotease.";
 RL J. Biochem. 122:438-452(1997).
 RN [7]
 RP ALTERNATIVE SPLICING (ISOFORM PACE4CS).
 RX MEDLINE=97064242; PubMed=8906861;
 RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
 RT "Functional analysis of human PACE4-A and PACE4-C isoforms:
 RT identification of a new PACE4-CS isoform.";
 RL FEBS Lett. 396:31-36(1996).
 RN [8]
 RP CHARACTERIZATION.
 RX MEDLINE=99233559; PubMed=10215603;
 RA Sucic J.F., Moehring J.M., Innocencio N.M., Luchini J.W.,
 RA Moehring T.J.;
 RT "Endoprotease PACE4 is Ca²⁺-dependent and temperature-sensitive and
 RT can partly rescue the phenotype of a furin-deficient cell strain.";
 RL Biochem. J. 339:639-647(1999).
 RN [9]
 RP PROCESSING.
 RX MEDLINE=98408849; PubMed=9738469;
 RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
 RA Tsuji A., Matsuda Y.;
 RT "Biosynthetic processing and quaternary interactions of propeptide
 RT convertase SPC4 (PACE4).";
 RL FEBS Lett. 434:155-159(1998).
 CC -!- FUNCTION: LIKELY TO REPRESENT AN ENDOPEPTIDASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS,
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -!- COFACTOR: PACE4A IS PROBABLY CALCIUM-DEPENDENT.
 CC -!- SUBUNIT: THE PACE4A-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE
 CC RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX
 CC WHEREAS MATURE PACE4A-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT
 CC PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.
 CC -!- SUBCELLULAR LOCATION: PACE4A-I AND PACE4A-II ARE SECRETED. PACE4C
 CC AND PACE4CS ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM
 CC IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED
 CC INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-
 CC TERMINUS. PACE4B MIGHT BE SECRETED.
 CC -!- ALTERNATIVE PRODUCTS: 8 ISOFORMS; PACE4A-I/PACE4 (SHOWN HERE),
 CC PACE4A-II, PACE4B/PACE4-I, PACE4C, PACE4CS, PACE4D, PACE4E-I AND
 CC PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B,
 CC C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.
 CC -!- TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE
 CC RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN,
 CC PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT
 CC COMPARATIVELY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST
 CC EXPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC
 CC KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE
 CC EXPRESSED IN PLACENTA. PACE4E-I IS EXPRESSED IN CEREBELLUM,
 CC PLACENTA AND PITUITARY. PACE4E-II IS AT LEAST PRESENT IN
 CC CEREBELLUM.
 CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM. ISOFORM PACE4D LACKS THE PROPEPTIDE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

Best Local Similarity 60.0%; Pred. No. 24; Mismatches 15; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 11 AARAGGGGGGIEGPTLRQCLAA 35
DB 24 AAGAGGAGGAGGAGGPGFRP-LAPR 47

RESULT 20
BET3_MESAU STANDARD; PRT; 367 AA.

AC 009029;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE BET3 protein.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96140430; PubMed=8552091;
RA Peyton M., Stellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,
RA Tsai M.-J.;
RT "BETA3, a novel helix-loop-helix protein, can act as a negative
RT regulator of BETA2 and Myob-responsive genes.";
RL Mol. Cell. Biol. 16:626-633(1996).
CC -!- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3
CC (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF
CC NEUROD1 AND MYOD-RESPONSIVE GENES. PROBABLY BY HETERODIMERIZATION
CC WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE
CC OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.
CC -!- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL: S80870; AAB50691.1; .
CC InterPro: IPR003015; HLH_MYC.
CC InterPro: IPR001092; HLH_dlm.
CC Pfam: PF00010; HLH; 1.
CC SMART: SM00353; HLH; 1.
CC PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
CC Nuclear protein; Transcription regulation; Repressor.
KW DOMAIN 11 14 POLY-ALA.
FT DOMAIN 58 62 POLY-SER.
FT DOMAIN 83 99 POLY-GLY.
FT DOMAIN 174 179 POLY-GLY.
FT DOMAIN 204 217 POLY-GLY.
FT DNA_BIND 229 240 BASIC DOMAIN.
FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 311 319 POLY-ALA.
SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AFF9E85F77 CRC64;

Query Match 31.1%; Score 59; DB 1; Length 367;
Best Local Similarity 61.1%; Pred. No. 12;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 11 AARAGGGGGGIEGPTLR 28
DB 88 AGGGGGGGGGGVSVPGL 105

Query Match 31.3%; Score 59.5; DB 1; Length 969;

```

RESULT 21
FXD2_HUMAN
ID FXD2_HUMAN STANDARD; PRT; 497 AA.
AC O60548;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-
DE related transcription factor 9) (FREAC-9).
GN FOXD2 OR FKHL17 OR FREAC9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98066765; PubMed=9403061;
RA Ernstsos S., Betz R., Lagercrantz S., Larsson C., Ericksson S.,
RA Cederberg A., Carlsson P., Enerbaeck S.;
RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
RL expressed human forkhead gene that maps to chromosome 1p32-p34.";
RN Genomics 46:78-85(1997).
RN [2]
RP REVISIONS.
RA Enerbaeck S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF042832; AAC15421.1;
DR HSP; O63245; 2FH.
DR TRANSFAC; T02485;
DR MIM; 602211;
DR InterPro: IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00339; FORK_HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
KW DOMAIN 90 94 POLY-ALA.
FT DOMAIN 101 104 POLY-ALA.
FT DNA_BIND 126 217 FORK-HEAD.
FT DOMAIN 247 250 POLY-ALA.
FT DOMAIN 296 306 POLY-ALA.
FT DOMAIN 398 409 POLY-GLY.
FT DOMAIN 421 426 POLY-GLY.
FT DOMAIN 442 445 POLY-ALA.
SQ SEQUENCE 497 AA; 49007 MW; EAAR498D216BE019 CRC64;

Query Match 31.1%; Score 59; DB 1; Length 497;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 4 PT--LRQCLAAAGGGGGGG 22
DB 385 PTALLRGLKTDAGGAGGG 405

RESULT 22
Y967_TREPA
ID Y967_TREPA STANDARD; PRT; 517 AA.
AC O83933;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TP0967.
GN TP0967.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
ON NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -!- SIMILARITY: BELONGS TO THE TP096X FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001264; AAC65925.1;
DR TIGR; TP0967;
KW Hypothetical protein; Complete proteome.
FT DOMAIN 152 161 POLY-GLY.
SQ SEQUENCE 517 AA; 55597 MW; E224976333989DF6 CRC64;

Query Match 31.1%; Score 59; DB 1; Length 517;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 GPTLRQCLAAAGGGGGGG 22
DB 141 GMTVTQPNAGAGGGGGGG 160

RESULT 23
KICJ_HUMAN
ID KICJ_HUMAN STANDARD; PRT; 593 AA.
AC P13645;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (CK 10).
GN KRT10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89125611; PubMed=2464696;
RA Rieger M., Franke W.W.;
RT "Identification of an orthologous mammalian cytokeratin gene. High
RT degree of intron sequence conservation during evolution of human
RT cytokeratin 10.";
RL J. Mol. Biol. 204:841-856(1988).
RN [2]
RP SEQUENCE OF 130-593 FROM N.A.
RX MEDLINE=88122104; PubMed=2448602;

```

Wed Oct 9 10:29:57 2002

RA Darmon M.Y., Semat A., Darmon M.C., Vasseur M.;
 RT "Sequence of a cDNA encoding human keratin No 10 selected according
 RT to structural homologies of keratins and their tissue-specific
 RT expression."; Mol. Biol. Rep. 12:277-283(1987).
 RL [3]
 RN SEQUENCE OF 197-593 FROM N.A.
 RP MEDLINE-92339897; PubMed-1378806;
 RX Tkachenko A.V., Buchman V.L., Blikskovsky V.V., Shvets Y.P.,
 RA Kisselev L.L.;
 RA "Exons I and VII of the gene (Ker10) encoding human keratin 10
 RT undergo structural rearrangements within repeats.";
 RT Gene 116:245-251(1992).
 RL [4]
 RN SEQUENCE OF 180-184 AND 577-589.
 RP TISSUE-Keratinocytes;
 RC MEDLINE-93162043; PubMed-1286667;
 RX Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
 RA Vandekerckhove J.;
 RA "Microsequences of 145 proteins recorded in the two-dimensional gel
 RT protein database of normal human epidermal keratinocytes.";
 RT Electrophoresis 13:960-969(1992).
 RL [5]
 RN VARIANT EHK HIS-156.
 RP MEDLINE-92386600; PubMed-1381287;
 RX Cheng J., Syder A.J., Yu Q.-C., Letai A., Paller A.S., Fuchs E.;
 RA "The genetic basis of epidermolytic hyperkeratosis: a disorder of
 RT differentiation-specific epidermal keratin genes.";
 RT Cell 70:811-819(1992).
 RL [6]
 RN VARIANTS.
 RP MEDLINE-92141228; PubMed-1371013;
 RX Korge B.P., Gan S.-Q., McBridge O.W., Mischke D., Steinert P.M.;
 RA "Extensive size polymorphism of the human keratin 10 chain resides in
 RT the C-terminal v2 subdomain due to variable numbers and sizes of
 RT glycine loops."; Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).
 RL [7]
 RN VARIANTS EHK HIS-156 AND SER-161.
 RP MEDLINE-92376531; PubMed-1380725;
 RX Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A., Roop D.R.;
 RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D.;
 RA "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
 RT hyperkeratosis."; Science 257:1128-1130(1992).
 RL [8]
 RN VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.
 RP MEDLINE-94136477; PubMed-7508181;
 RX Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,
 RA Compton J.G., Bale S.J.;
 RA "Preferential sites in keratin 10 that are mutated in epidermolytic
 RT hyperkeratosis."; Am. J. Hum. Genet. 54:179-190(1994).
 RL [9]
 RN VARIANTS EHK ARG-150; CYS-156 AND GLU-439, AND VARIANT SER-126.
 RP MEDLINE-94216497; PubMed-7512983;
 RX Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;
 RA "Genetic mutations in the K1 and K10 genes of patients with
 RT epidermolytic hyperkeratosis. Correlation between location and
 RT disease severity."; J. Clin. Invest. 93:1533-1542(1994).
 RL [10]
 RN VARIANT EHK ASN-160.
 RP MEDLINE-94117868; PubMed-7507150;
 RX Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;
 RA "Prenatal diagnosis of epidermolytic hyperkeratosis by direct gene
 RT sequencing."; J. Invest. Dermatol. 102:13-16(1994).
 RL [11]
 RN VARIANTS EHK PRO-156 AND SER-156.
 RP MEDLINE-94117870; PubMed-7507152;
 RX McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,
 RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
 RA Morley S.M.;
 RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
 RT congenital ichthyosiform erythroderma (BCIE).";
 RT J. Invest. Dermatol. 102:24-30(1994).
 RL [12]
 RN VARIANT EHK THR-150.
 RP MEDLINE-95059228; PubMed-7526210;
 RX Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G.,
 RA Fuchs E.;
 RA "Genetic and clinical mosaicism in a type of epidermal nevus.";
 RT New Engl. J. Med. 331:1408-1415(1994).
 RL [13]
 RN VARIANT AEI THR-446.
 RP MEDLINE-99072665; PubMed-9856845;
 RX Suga Y., Duncan K.O., Heald P.W., Roop D.R.;
 RA "A novel helix termination mutation in keratin 10 in annular
 RT epidermolytic ichthyosis, a variant of bullous congenital
 RT ichthyosiform erythroderma.";
 RT J. Invest. Dermatol. 111:1220-1223(1998).
 RL [14]
 RN VARIANT EHK SER-160.
 RP MEDLINE-99215719; PubMed-10201536;
 RX Arin M.J., Longley M.A., Anton-Lamprecht I., Kurze G., Huber M.,
 RA Hohl D., Rothnagel J.A., Roop D.R.;
 RA "A novel substitution in keratin 10 in epidermolytic hyperkeratosis.";
 RT Invest. Dermatol. 112:506-508(1999).
 RL [15]
 RN SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
 CC -1- KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1
 CC -1- TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING
 CC STRATUM CORNEUM.
 CC -1- POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN
 CC THE GLY-RICH REGION (POSITIONS 490-560).
 CC -1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF EPIDERMOLYTIC
 CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
 CC ICHTHYOSIFORM ERYTHRODERMA (BCIE)); A HEREDITARY SKIN DISORDER
 CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
 CC CORNEUM. AT BIRTH, AFFECTED INDIVIDUALS USUALLY PRESENT WITH
 CC REDNESS, BLISTERS AND SUPERFICIAL EROSIONS DUE TO CYTOLYSIS.
 CC WITHIN A FEW WEEKS, THE ERYTHRODERMA AND BLISTER FORMATION
 CC DIMINISH AND HYPERKERATOSES DEVELOP. TRANSMISSION IS AUTOSOMAL
 CC DOMINANT, BUT MOST CASES ARE SPORADIC.
 CC -1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF ANNULAR EPIDERMOLYTIC
 CC ICHTHYOSIS (AEI), A DISTINCT PHENOTYPIC VARIANT OF EPIDERMOLYTIC
 CC HYPERKERATOSIS. IT RESEMBLES CLINICAL AND HISTOLOGIC FEATURES OF
 CC BOTH EPIDERMOLYTIC HYPERKERATOSIS AND ICHTHYOSIS BULLOSA OF
 CC STEVENS.
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
 CC MICROFILAMENTAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II
 CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN
 CC POSITIONS 513 TO 555.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X14487; CAA32649.1; -
 CC EMBL; M19156; AAA59468.1; -
 CC EMBL; M77663; AAA59199.1; -
 CC EMBL; L20218; AAB59438.1; -
 CC EMBL; L20219; AAB59439.1; -
 CC PIR; S02158; KRHU0.
 CC Aarhus/Ghent-2DPAGE; 7405; IEF.
 CC MIM; 148080; -
 CC MIM; 113800; -
 CC InterPro: IPR001664; IF.
 CC InterPro: IPR002957; Keratin_1.
 CC Pfam; PF00038; filament; 1.


```

DR PRINTS: PRO1248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin; Disease mutation;
KW Polymorphism.
FT DOMAIN 1 145 HEAD.
FT DOMAIN 146 456 ROD.
FT DOMAIN 457 593 TAIL.
FT DOMAIN 146 181 COIL 1A.
FT DOMAIN 182 202 LINKER 1.
FT DOMAIN 203 294 COIL 1B.
FT DOMAIN 295 317 LINKER 12.
FT DOMAIN 318 456 COIL 2.
FT DOMAIN 457 590 GLY/PHE/SER-RICH.
FT VARIANT 126 126 GLY/SER-RICH.
FT VARIANT 150 150 /FTID=VAR_010505.
FT VARIANT 150 150 M -> R (IN EHK).
FT VARIANT 150 150 /FTID=VAR_010506.
FT VARIANT 154 154 M -> T (IN EHK).
FT VARIANT 156 156 /FTID=VAR_010507.
FT VARIANT 156 156 N -> H (IN EHK).
FT VARIANT 156 156 /FTID=VAR_003826.
FT VARIANT 156 156 R -> H (IN EHK).
FT VARIANT 156 156 /FTID=VAR_003827.
FT VARIANT 156 156 R -> C (IN EHK).
FT VARIANT 156 156 /FTID=VAR_003828.
FT VARIANT 156 156 R -> P (IN EHK).
FT VARIANT 156 156 /FTID=VAR_003829.
FT VARIANT 156 156 R -> S (IN EHK).
FT VARIANT 160 160 /FTID=VAR_003830.
FT VARIANT 160 160 Y -> D (IN EHK; SEVERE PHENOTYPE).
FT VARIANT 160 160 /FTID=VAR_003831.

Query Match 31.1%; Score 59; DB 1; Length 593;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 11 ARAGGGGGGGGIEG 25
Db 13 SSRSGGGGGGGCGG 27

RESULT 24
ZIN_HUMAN STANDARD; PRT; 753 AA.
AC Q9NRL3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Zinedin.
GN ZIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347911; PubMed=10748158;
RA Castets F., Rakitina T., Gaillard S., Moqrich A., Mattei M.-G.,
RA Monneron A.;
RT "Zinedin, SG2NA, and striatin are calmodulin-binding, WD repeat
RT proteins principally expressed in the brain.";
RL J. Biol. Chem. 275:19970-19977(2000).
RN [2]
RP SEQUENCE OF 402-753 FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER. MAY
CC -!- FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.

```

```

CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -!- CAUTION: The name "Zinedin" probably originates from the name of
CC the famous soccer player from Marseille (Zinedine Zidane)!
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF212940; AAF29527.1; -
CC EMBL; BC004910; AAH04910.1; -
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00320; GPROTEINRPT.
CC SMART; SM00320; WD40; 6.
CC PROSITE; PS00678; WD_REPEATS_1; 1.
CC PROSITE; PS0082; WD_REPEATS_2; 4.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Calmodulin-binding; Repeat; WD repeat; Coiled coil.
FT DOMAIN 69 136 COILED COIL (POTENTIAL).
FT DOMAIN 165 182 CALMODULIN-BINDING (POTENTIAL).
FT REPEAT 436 475 WD 1.
FT REPEAT 489 528 WD 2.
FT REPEAT 542 581 WD 3.
FT REPEAT 587 628 WD 4.
FT REPEAT 635 674 WD 5.
FT REPEAT 677 716 WD 6.
FT REPEAT 723 752 WD 7.
FT SITE 71 79 CAVEOLIN-BINDING (POTENTIAL).
FT DOMAIN 6 14 POLY-ALA.
FT CONFLICT 402 404 LAD -> GTR (IN REF. 2).
SQ SEQUENCE 753 AA; 4DA016A8FF7EDB5E CRC64;

Query Match 31.1%; Score 59; DB 1; Length 753;
Best Local Similarity 78.6%; Pred. No. 21;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 AGGGGGGGGIEGPT 27
Db 44 AGGGGGGGGSPGPT 57

RESULT 25
ECR_LUCCU STANDARD; PRT; 757 AA.
AC O18531;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone
DE receptor) (20E receptor).
GN ECR OR NR1H1.
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Oestroidea; Calliphoridae; Luciflia.
OX NCBI_TaxID=7375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97449774; PubMed=9304790;
RA Hannan G.N., Hill R.J.;
RT "Cloning and characterization of LcEcR: a functional ecdysone
RT receptor from the sheep blowfly Lucilia cuprina.";
RL Insect Biochem. Mol. Biol. 27:479-488(1997).
CC -!- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
CC ELEMENTS (ECRES) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC -!- SIMILARITY: NR1 SUBFAMILY.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75355; AAB81130.1; -
CC HSP; P20393; IA6Y.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR001723; Strdhormone_receptor.
CC InterPro; IPR001628; zf-C4.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; ZNF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
FT DOMAIN 1 300 MODULATING (POTENTIAL).
FT DNA_BIND 301 366 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 301 321 C4-TYPE.
FT ZN_FING 337 361 C4-TYPE.
FT DOMAIN 454 674 HORMONE-BINDING (POTENTIAL).
SQ SEQUENCE 757 AA; 83075 MW; C1511452ED37D359 CRC64;

Query Match 31.1%; Score 59; DB 1; Length 757;
Best Local Similarity 76.9%; Pred. No. 21;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27
DB 129 GGGGGGGGVPGMT 141

RESULT 26
SSB_MYCLE
ID SSB_MYCLE STANDARD; PRT; 168 AA.
AC P46390; O53126;
DT 01-NOV-1995 (Rel. 32; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Single-strand binding protein (SSB) (Helix-destabilizing protein).
GN SSB OR ML2684 OR MLCB1913.20C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97124199; PubMed=8969512;
RA Fsihl H., de Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
RA Takiff H.E., Eiglmeyer K., Bergh S., Cole S.T.;
RT "Gene arrangement and organization in a approximately 76 kb fragment
RT encompassing the oric region of the chromosome of Mycobacterium
RT leprae".
RL Microbiology 142:3147-3161(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
CC CHROMOSOME. IT IS ALSO INVOLVED IN DNA RECOMBINATION AND REPAIR
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SSB FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 137.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39923; AAB53120.1; ALT_FRAME.
DR EMBL; AL022118; CAA17953.1; -
DR EMBL; AL583926; CAC32216.1; -
DR Leptoma; ML2684; -
DR HSP; P02339; LEYG.
DR InterPro; IPR000424; SSB.
DR Pfam; PF00436; SSB; 1.
DR PROSITE; PS00735; SSB_1; FALSE_NEG.
DR PROSITE; PS00736; SSB_2; FALSE_NEG.
KW DNA-binding; DNA repair; DNA replication; Complete proteome.
FT DOMAIN 124 133 POLY-GLY.
SQ SEQUENCE 168 AA; 17700 MW; 077C62E430623658 CRC64;

Query Match 30.8%; Score 58.5; DB 1; Length 168;
Best Local Similarity 52.0%; Pred. No. 6.7;
Matches 13; Conservative 3; Mismatches 4; Indels 5; Gaps 1;

QY 3 GPTLRQCL-----AARAGGGGGGG 22
DB 107 GSLRYATAKVNKASRGSGGGGFG 131

RESULT 27
HH3R_HUMAN
ID HH3R_HUMAN STANDARD; PRT; 445 AA.
AC Q9Y5N1; Q9H4K8; Q9G2X2;
DT 01-MAR-2002 (Rel. 41; Created)
DT 01-MAR-2002 (Rel. 41; Last sequence update)
DT 01-MAR-2002 (Rel. 41; Last annotation update)
DE Histamine H3 receptor (HH3R) (G protein-coupled receptor 97).
GN HRH3 OR GPCR97.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Thalamus;
RX MEDLINE=99278519; PubMed=10347254;
RA Lovenberg T.W., Roland B.L., Wilson S.J., Jiang X., Pyati J.,
RA Huvar A., Jackson M.R., Erlender M.G.;
RT "Cloning and functional expression of the human histamine H3
RT receptor".
RL Mol. Pharmacol. 55:1101-1107(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RX MEDLINE=20568725; PubMed=11118334;
RA Nakamura T., Itadani H., Hidaka Y., Ohta M., Tanaka K.;
RT "Molecular cloning and characterization of a new human histamine
RT receptor, HH4R".
RL Biochem. Biophys. Res. Commun. 279:615-620(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4; 5; 6 AND 7).
RX TISSUE=Thalamus;
RX MEDLINE=21181559; PubMed=11284713;
RA Coe F., Guenin S.-P., Audinot V., Renouard-Try A., Beauverger P.,

RA Macia C., Ouvre C., Nagel N., Rique H., Boutin J.A., Galizzi J.-P.;
 RT "Genomic organization and characterization of splice variants of the
 RL human histamine H3 receptor";
 RN Biochem. J. 355:279-288(2001).
 (4)

RC SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT VAL-280.
 RD TISSUE-Blood,
 RE Wiedemann P., Bonisch H., Bruss M.;

RA "An amino acid variation in the human histamine h3 receptor from a
 RT patient suffering from orthostatic dysregulation";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 (5)

RC SEQUENCE FROM N.A. (ISOFORM 3).
 RD Ullmer C., Zirves E., Lubbert H.;

RA "Cloning and functional expression of the human histamine H3
 RT receptor";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 (6)

RC SEQUENCE FROM N.A. (ISOFORM 1).
 RD Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhani P., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Levasaiho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McMay K., McMurray A.A.,

RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showknet R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20";
 RL Nature 414:865-871(2001).

CC -1- FUNCTION: THE H3 SUBCLASS OF HISTAMINE RECEPTORS COULD MEDIATE THE
 CC HISTAMINE SIGNALS IN CNS AND PERIPHERAL NERVOUS SYSTEM. SIGNALS

CC THROUGH THE INHIBITION OF ADENYLATE CYCLASE AND DISPLAYS HIGH
 CC CONSTITUTIVE ACTIVITY (SPONTANEOUS ACTIVITY IN THE ABSENCE OF

CC AGONIST). AGONIST STIMULATION OF ISOFORM 3 NIETHER MODIFIED
 CC ADENYLATE CYCLASE ACTIVITY NOR INDUCED INTRACELLULAR CALCIUM

CC MOBILIZATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; 1 (SHOWN HERE), 2;
 CC 3/H3S; 4; 5; 6 AND 7; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE CNS, WITH THE
 CC GREATEST EXPRESSION IN THE THALAMUS AND CAUDATE NUCLEUS. THE

CC VARIOUS ISOFORMS ARE MAINLY COEXPRESSED IN BRAIN, BUT THEIR
 CC RELATIVE EXPRESSION LEVEL VARIES IN A REGION-SPECIFIC MANNER.

CC ISOFORMS 3 AND 7 ARE HIGHLY EXPRESSED IN THE THALAMUS, CAUDATE
 CC NUCLEUS AND CEREBELLUM WHILE ISOFORMS 5 AND 6 SHOW A POOR

CC EXPRESSION. ISOFORMS 5 AND 6 SHOW A HIGH EXPRESSION IN THE
 CC AMYGDALA, SUBSTANTIA NIGRA, CEREBRAL CORTEX AND HYPOTHALAMUS.

CC ISOFORM 7 IS NOT FOUND IN HYPOTHALAMUS OR SUBSTANTIA NIGRA.
 CC MISCELLANEOUS: Does not bind to cimetidine and triptolide. Shows

CC modest affinity for thioperamide. Isoform 4 is unable to bind to
 CC and R(-)-alpha-methylhistamine. Isoform 4 is unable to bind to

CC iodoproxyfan while isoforms 1 and 3 bind it with high affinity.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF140538; AAD38151.1; -
 DR EMBL; AB045369; BAB20090.1; -
 DR EMBL; AB019000; BAB17030.1; -
 DR EMBL; AJ296652; CAC51025.1; -
 DR EMBL; AJ278250; CAC39434.1; -
 DR EMBL; AL078633; CAC04014.1; -
 DR EMBL; AF363791; AAK50040.1; -
 DR MIM; 604525; -
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm1.1;
 DR PRINTS; PR00237; GPCRHOOPS

DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1;
 DR PROSITE; PS00237; G_PROTEIN_RECP_F1_2; 1;
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Alternative splicing; Disease mutation.

FT DOMAIN 1 39 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 40 60 POTENTIAL.
 FT DOMAIN 61 70 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 71 91 POTENTIAL.
 FT DOMAIN 92 108 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 109 129 POTENTIAL.
 FT DOMAIN 130 156 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 157 177 POTENTIAL.
 FT DOMAIN 178 196 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 197 217 POTENTIAL.
 FT DOMAIN 218 359 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 360 380 POTENTIAL.
 FT DOMAIN 381 395 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 396 416 POTENTIAL.
 FT DOMAIN 417 445 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 446 466 POLY-ALA.
 FT DOMAIN 250 256 POLY-PRO.
 FT TRANSMEM 292 298 POLY-GLY.
 FT CARBOHYD 11 11 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 15 98 MISSING (IN ISOFORM 4).
 FT VARSPLIC 197 315 MISSING (IN ISOFORM 5).
 FT VARSPLIC 227 342 MISSING (IN ISOFORM 6).
 FT VARSPLIC 234 263 MISSING (IN ISOFORM 7).
 FT VARSPLIC 274 353 MISSING (IN ISOFORM 3).
 FT VARSPLIC 445 445
 FT VARIANT 280
 FT CONFLICT 19 19
 FT SEQUENCE 445 AA; 48671 MW; 2ACF7440FBE95B6C CRC64;

Query Match 30.5%; Score 58; DB 1; Length 445;
 Best Local Similarity 62.5%; Pred. No. 17;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 12 ARAGGGGGGGIEGPT 27
 Db 289 ATLGGGGGGGSVASPT 304
 I | | | | | | | | | |

RESULT 28
 EVX2_HUMAN STANDARD; PRT; 476 AA.

AC Q03828;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Homeobox even-skipped homolog protein 2 (EVX-2).
 GN EVX2.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

```

[1]
RN      SEQUENCE FROM N.A.
RP      Birren B., Linton L., Nusbaum C., Lander E.;
RA      Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN      SEQUENCE OF 144-300 FROM N.A.
RP      MEDLINE=91257849; PubMed=1675198;
RA      D'Esposito M., Morelli F., Acampora D., Migliaccio E., Simeone A.,
RT      Boncinelli E.;
RT      "EVX2", a human homeobox gene homologous to the even-skipped
RT      segmentation gene, is localized at the 5' end of HOX4 locus on
RT      chromosome 2.;
RT      Genomics 10:43-50(1991).
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- DEVELOPMENTAL STAGE: EXPRESSED DURING EARLY EMBRYOGENESIS AND
CC      NEUROGENESIS IN A BIPHASIC MANNER.
CC      -!- SIMILARITY: BELONGS TO THE EVEN-SKIPPED FAMILY OF HOMEBOX
CC      PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
EMBL; AC009336; -; NOT_ANNOTATED_CDS.
DR      EMBL; M59983; AAA52414.1; -;
DR      EMBL; M59982; AAA52414.1; JOINED.
DR      HSSP; P14653; I872.
DR      MIM; 142991; -;
DR      InterPro; IPR000047; HTH_repressor.
DR      InterPro; IPR001356; Homeobox.
DR      Pfam; PF000046; homeobox; 2.
DR      PRINTS; PR000024; HOMEBOX.
DR      PRINTS; PR000031; HTHREPRESSR.
DR      PRINTS; SM00389; HOX; 1.
DR      SMART; PS00021; HOMEBOX_1; 1.
DR      PROSITE; PS50071; HOMEBOX_2; 1.
KW      DNA-binding; Developmental protein; Homeobox; Nuclear protein.
FT      DNA_BIND 188 247
FT      DOMAIN 294 301 POLY-ALA.
FT      DOMAIN 304 308 POLY-ALA.
FT      DOMAIN 346 351 POLY-ALA.
FT      DOMAIN 356 370 POLY-ALA.
FT      DOMAIN 373 378 POLY-ALA.
FT      DOMAIN 398 408 POLY-ALA.
FT      DOMAIN 413 434 POLY-GLX.
SQ      SEQUENCE 476 AA; 47799 MW; 6AA99041BA151C3F CRC64;

Query Match 30.5%; Score 58; DB 1; Length 476;
Best Local Similarity 76.9%; Pred. No. 18;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGG 22
Db 409 LGSRRGGGGGGG 421
| : | | | | | | | |
| : | | | | | | | |

RESULT 29
ONC2_HUMAN
ID ONC2_HUMAN STANDARD; PRT; 485 AA.
AC O95948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE One cut domain family member 2 (ONECUT-2 transcription factor) (OC-2).
GN ONECUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;

```

```
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
CC TO CLASS-3 POU.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M82999; AAA39960.1; -
CC PIR; S31223; S31223.
CC HSP; P14859; IOCT.
CC MGD; MGI:102564; Pou3f3.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR000327; POU.
CC Pfam; PF00046; homeobox; 1.
CC Pfam; PF00157; pou; 1.
CC PRINTS; PR00028; POUDOMAIN.
CC PRODom; PD000583; POU; 1.
CC SMART; SM00389; Hox; 1.
CC SMART; SM00352; POU; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS50071; HOMEBOX_2; 1.
CC PROSITE; PS00035; POU_1; 1.
CC PROSITE; PS00465; POU_2; 1.
CC Nuclear protein; DNA-binding; Homeobox.
KW DOMAIN 28 49 POLY-GLY.
FT DOMAIN 101 112 POLY-ALA.
FT DOMAIN 186 201 POLY-ALA.
FT DOMAIN 267 291 HIS-RICH.
FT DOMAIN 313 383 POU.
FT DNA_BIND 401 460 HOMEBOX.
SQ SEQUENCE 495 AA; 50012 MW; 77B802E890C9A014 CRC64;

Query Match 30.5%; Score 58; DB 1; Length 495;
Best Local Similarity 91.7%; Pred. No. 19;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGG 22
Db 26 AAGAGGGGGGGG 37
```

Search completed: October 9, 2002, 09:00:15
Job time : 5.3831 secs

11/11/11

us-09-422-838c-28.rspt

Wed Oct 9 10:29:58 2002

```

AC Q943K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P0039A07.6 PROTEIN.
GN P0039A07.6
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OC NCBI_TaxID=4530;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0039A07.6";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003235; BAB64100.1; -.
SQ SEQUENCE 253 AA; 25368 MW; A563166CE5F97B2B CRC64;

Query Match 37.4%; Score 71; DB 10; Length 253;
Best Local Similarity 55.6%; Pred. No. 0.86;
Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 3 GPTLRQCLAAAGGGGGGGGIEGPTLR 29
||| : ||| ||||| | |
DB 80 GPTVGVVAVRAGAGGGGGGPRGFAK 106

RESULT 3
Q9LYB2 PRELIMINARY; PRT; 199 AA.
ID Q9LYB2
AC Q9LYB2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 21.5 KDA PROTEIN.
DR T20010.200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RA Obermaier B., Ottenwaelder B., Duchemin D., Zeitler K., Mewes H.W.,
RU R. S. Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163816; CAB87755.1; -.
DR InterPro: IPR000345; Cyt_c_heme_bind.
DR InterPro: IPR002395; Kininogen.
DR PRINTS: PR00334; KININOGEN.
DR PROSITE: PS00130; CYTOCHROME_C; UNKNOWN_1.
KW Hypothetical protein
SQ SEQUENCE 199 AA; 21539 MW; E5D28AC167B3FBF8 CRC64;

Query Match 35.8%; Score 68; DB 10; Length 199;
Best Local Similarity 34.8%; Pred. No. 1.5;
Matches 16; Conservative 3; Mismatches 11; Indels 16; Gaps 1;

QY 2 EGPTLRQC-----LAARAGGGGGGGGIEGPTLRQC 31
||| : ||| ||||| : |
DB 7 EGRTRCPFASTCTCTLTVAQTSLLCVDDGGGGGGGVDGVDRC 52

RESULT 4
Q9P270 PRELIMINARY; PRT; 612 AA.
ID Q9P270

```

```

AC Q9P270;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA1458 PROTEIN (FRAGMENT).
GN KIAA1458.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
genes.XVII.The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040891; BAA95982.1; -.
FT NON_TER 1
SQ SEQUENCE 612 AA; 65593 MW; 9AA4061D21E1E9FD CRC64;

Query Match 35.8%; Score 68; DB 4; Length 612;
Best Local Similarity 63.6%; Pred. No. 4.4;
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLRQCLAAAGGGGGGGGIEG 25
||| : ||| ||||| | |
DB 10 PSLSLRLRRAGGGGGGGGAG 31

RESULT 5
Q9IBC5 PRELIMINARY; PRT; 66 AA.
ID Q9IBC5
AC Q9IBC5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 7.0 KDA PROTEIN.
OS Spodoptera litura nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OC NCBI_TaxID=46242;
RN
RP SEQUENCE FROM N.A.
RA STRAIN=G2;
RX MEDLINE=21425398; PubMed=115314116;
RA Pang Y., Yu J., Wang L., Hu X., Bao W., Li G., Chen C., Han H., Hu S.,
RA Yang H.;
RT "Sequence Analysis of the Spodoptera litura Multicapsid
Nucleopolyhedrovirus Genome.";
RL Virology 287:391-404(2001).
RN
RP SEQUENCE FROM N.A.
RA STRAIN=G2;
RC Yu J., Wang L., Hu X., Pang Y.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325155; RAL01786.1; -.
KW Hypothetical protein
SQ SEQUENCE 66 AA; 6998 MW; C5626A8FFA9C9E7C CRC64;

Query Match 34.7%; Score 66; DB 12; Length 66;
Best Local Similarity 54.5%; Pred. No. 0.89;
Matches 12; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 7 RQCLAAAGGGGGGGGIEGPTL 28
||| : ||| ||||| : |
DB 13 QOASSNRSGGGGGGGGVVGMAL 34

RESULT 6
Q9M6A1 PRELIMINARY; PRT; 137 AA.
ID Q9M6A1

```


AC Q9M6A1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE GLYCINE-RICH RNA BINDING PROTEIN 1.
 GN GRP-1
 OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 CC Vinales; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Clastre M., Chenieux J.-C., Rideau M., Hamdi S.;
 RT "Genes encoding glycine-rich Catharanthus roseus proteins with RNA-
 binding motifs."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF200321; AAF31402.1; 1.
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; RRM; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP.1; 1.
 DR PROSITE; PS00030; RRM_RNP.1; 1.
 SQ SEQUENCE 137 AA; 14162 MW; 4FABADB9C7A989FC CRC64;
 Query Match 34.7%; Score 66; DB 10; Length 137;
 Best Local Similarity 50.0%; Pred. No. 1.8;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 QY 5 TLRCCLAAAGGGGGGIEGP 26
 Db 80 TVNEAQRSGGGGGGFRGP 101
 RESULT 7
 Q9M699
 ID Q9M699 PRELIMINARY; PRT; 160 AA.
 AC Q9M699;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE GLYCINE-RICH RNA-BINDING PROTEIN 2.
 GN GRP-2.
 OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 CC Vinales; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Courtois M., Chenieux J.-C., Hamdi S., Rideau M.,
 RA Clastre M.;
 RT "Cloning of two cDNAs encoding cRGP2 and cRGP3 (Accession Nos.
 AF200323 and AF200322), the first members of the RRM-GRP family in
 Catharanthus roseus (PGR00-049)."
 RL Plant Physiol. 122:1459-1459(2000).
 DR EMBL; AF200323; AAF31404.1; 1.
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; RRM; 1.
 DR PRINTS; PR01228; EGGSHELL.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP.1; 1.
 SQ SEQUENCE 160 AA; 16264 MW; DDC9F63C983F5F2 CRC64;
 Query Match 34.7%; Score 66; DB 10; Length 160;
 Best Local Similarity 50.0%; Pred. No. 2.1;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLRCCLAAAGGGGGGIEGP 26
 Db 80 TVNEAQRSGGGGGGFRGP 101
 RESULT 8
 Q9XE89
 ID Q9XE89 PRELIMINARY; PRT; 369 AA.
 AC Q9XE89;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 39.1 KDA PROTEIN.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 CC Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Llaca V., Lou A., Messing J.W.;
 RT "Microsytendy analysis of 22-kda zein cluster in maize and sorghum."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF061282; AAD22156.1; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 369 AA; 39080 MW; DAA3C65088F106CE CRC64;
 Query Match 34.7%; Score 66; DB 10; Length 369;
 Best Local Similarity 40.0%; Pred. No. 4.5;
 Matches 14; Conservative 7; Mismatches 12; Indels 2; Gaps 1;
 QY 4 PTLRQCLAAAGGGGGGIEPTLRQ--CLAARA 36
 Db 167 PAKKASISASVGGGGGGGWWRRGPGCSGRS 201

RESULT 9
 Q9LI26
 ID Q9LI26 PRELIMINARY; PRT; 413 AA.
 AC Q9LI26;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhartioideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone:PO708G02."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP001539; BAA92912.1; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 413 AA; 45035 MW; 4FEC2A4C5D1271CF CRC64;
 Query Match 34.7%; Score 66; DB 10; Length 413;
 Best Local Similarity 64.7%; Pred. No. 5;
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 7 RCLAAARAGGGGGG 23
 Db 162 RRCAGLLAGGGGGGV 178
 RESULT 10
 Q96SQ2
 ID Q96SQ2 PRELIMINARY; PRT; 474 AA.

Wed Oct 9 10:29:58 2002

```

AC Q96S02:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA FLJ14713 FIS, CLONE NT2RP300845, MODERATELY SIMILAR TO
DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE P78 (EC 2.7.1.-).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027619; BAB5238.1; -
SQ SEQUENCE 474 AA; 51313 MW; 0BCA301518F20DED CRC64;

Query Match 34.7%; Score 66; DB 4; Length 474;
Best Local Similarity 45.0%; Pred. No. 5.8;
Matches 18; Conservative 1; Mismatches 15; Indels 6; Gaps 1;

OY 3 GPTLR-----QCLARAGGGGGGGGGGPTLRQCLARA 36
DB 284 GSTIRSFHGGQVRDRAGGGGGGGVQNGPPASPTLAHEA 323

RESULT 11
Q9AD76 PRELIMINARY; PRT; 496 AA.
AC Q9AD76:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN SKL13.27.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RC Seeger K.J., Harris D.;
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RC Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL512667; CAC21636.2; -
DR InterPro: IPR003838; DUF214.
DR Pfam: PF02687; DUF214; 1.
SQ SEQUENCE 496 AA; 49548 MW; 54E110C4F86231A4 CRC64;

Query Match 34.7%; Score 66; DB 2; Length 496;
Best Local Similarity 46.2%; Pred. No. 6;
Matches 18; Conservative 2; Mismatches 9; Indels 10; Gaps 2;

```

```

OY 4 PTLRQCLARAGGG-----GGGEGPTLRQCLA 33
DB 408 PTLQALCGGAGGGAGGGGGGGGGGGGGPG-ROAAA 445

RESULT 12
Q19476 PRELIMINARY; PRT; 500 AA.
AC Q19476:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F15B9.5 PROTEIN.
GN F15B9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z78013; CAB01420.1; -
DR InterPro: IPR001254; Trypsin.
DR PROSITE: P50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 500 AA; 53946 MW; 1416327086FE7CF6 CRC64;

Query Match 34.7%; Score 66; DB 5; Length 500;
Best Local Similarity 56.5%; Pred. No. 6.1;
Matches 13; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 3 GPTLRQCLARAGGGGGGGGGGGG 25
DB 429 GSMLGRFLSNRGGGGGGGGGG 451

RESULT 13
Q9BYD8 PRELIMINARY; PRT; 688 AA.
AC Q9BYD8:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAP/MICROTUBULE AFFINITY-REGULATING KINASE LIKE 1.
GN MARKL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=21226021; PubMed=11326310;
RA Kato T., Satoh S., Okabe H., Kitahara O., Ono K., Kihara C.,
RA Tanaka T., Tsunoda T., Yamaoka Y., Nakamura Y., Furukawa Y.;
RT "Isolation of a novel human gene, MARKL1, homologous to MARK3 and its
RT involvement in hepatocellular carcinogenesis.";
RL Neoplasia 3:4-9(2001).
CC -1; SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AB049127; BAB39380.1; -
DR HSSP: Q63450; 1A06.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

Wed Oct 9 10:29:58 2002

```

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=99263496; PubMed=10329000;
Heus H.C., Hing A., Baren M.Jvan, Joosse M., Breedveld G., Wang J.C.,
Burgess A., Donnis-Keller H., Berglund C., Scherer S.W., Rommens J.M.,
Oostra B.A., Heutink P.;
RA "A physical and transcriptional map of the preaxial polydactyly locus
RT on chromosome 7q36."
RL Genomics 57:342-351(1999).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: WITH OTHER HOMEBOX PROTEINS.
DR EMBL; AF107453; AAD41467.1; JOINED.
DR HSSP; P14653; 1B72.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_repressr.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
KW NON_TER 355
FT SEQUENCE 355 AA; 35587 MW; CD41D18CC811F0E9 CRC64;
SQ
Query Match 34.5%; Score 65.5; DB 4; Length 355;
Best Local Similarity 57.1%; Pred. No. 5;
Matches 16; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

Qy 10 LAARA---GGGGGGGGTGTGPTLRQCLAA 34
| | | | | | | | | | | | | | | | | |
Db 34 LAARAGTGGGGGGGGASGTGSCSPA 61
| | | | | | | | | | | | | | | | | |

RESULT 18
Q9GNP1 PRELIMINARY; PRT; 770 AA.
AC Q9GNP1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VASA HOMOLOG.
GN CSDEAD1B(CSVHB).
OS Ciona savignyi.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC Clonidae; Ciona.
OX NCBI_TaxID=51511;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OVARY;
RX MEDLINE=20130953; PubMed=10664149;
RA Fujimura M., Takamura K.;
RT "Characterization of an ascidian DEAD-box gene, Ci-DEAD1: specific
RT expression in the germ cells and its mRNA localization in the
RT posterior-most blastomeres in early embryos."
RL Dev. Genes Evol. 210:64-72(2000).
DR EMBL; AB047803; BAB12217.1; -.
DR HSSP; Q58083; 1HV8.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00098; zf-CCHC; 6.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00343; Znf_C2HC; 6.
KW ATP-binding; Helicase; Zinc-finger.

SQ SEQUENCE 381 AA; 40761 MW; F3A0E3CEBD950778 CRC64;
Query Match 34.2%; Score 65; DB 10; Length 381;
Best Local Similarity 48.1%; Pred. No. 6.1;
Matches 13; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 4 PTLRQCLAAAGGGGGGGTGTGPTLRQ 30
| | | | | | | | | | | | | |
Db 23 PHARRALALRVGGGGGPAFASLTAVRE 49
| | | | | | | | | | | | | |

RESULT 20
Q9VJK4 PRELIMINARY; PRT; 452 AA.
AC Q9VJK4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE CGS953 PROTEIN.
GN CGS953.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```

```

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman M.R., Bouck J., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boutin L., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Delcher A., Deng Z., Davenport L.B., Davies P.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Mays A.D., Dietz S.M.,
RA Folsler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodek A., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kallush F., Karpman C., Kravitz S., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AF003592; AAF51656.1; -
DR FLYBASE: FBgn0037034; CG13257.
DR InterPro: IPR001545; Glyco_hormone_beta.
DR Pfam: PF00057; ldl_recept_a; 1.
DR SMART: SM00192; LDLa; 1.
DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; UNKNOWN_1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 1.
KW Glycoprotein.
SQ SEQUENCE 452 AA; 47875 MW; 0F7ABD53014E3E5C CRC64;

Query Match 34.2%; Score 65; DB 5; Length 452;
Best Local Similarity 73.3%; Pred. No. 7.1;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 12 ARAGGGGGGGIRGP 26
Db 259 AAGGGGGGGGVVGP 273

RESULT 21
Q9VP99 ID Q9VP99 PRELIMINARY; PRT; 150 AA.
AC Q9VP99;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN CG13257 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman M.R., Bouck J., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boutin L., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Delcher A., Deng Z., Davenport L.B., Davies P.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Mays A.D., Dietz S.M.,
RA Folsler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodek A., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kallush F., Karpman C., Kravitz S., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AF003592; AAF53541.1; -
DR FLYBASE: FBgn0032587; CG5953.
DR InterPro: IPR001545; Glyco_hormone_beta.
DR Pfam: PF00057; ldl_recept_a; 1.
DR SMART: SM00192; LDLa; 1.
DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; UNKNOWN_1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 1.
KW Glycoprotein.
SQ SEQUENCE 452 AA; 47875 MW; 0F7ABD53014E3E5C CRC64;

Query Match 34.2%; Score 65; DB 5; Length 452;
Best Local Similarity 73.3%; Pred. No. 7.1;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 12 ARAGGGGGGGIRGP 26
Db 259 AAGGGGGGGGVVGP 273

RESULT 21
Q9VP99 ID Q9VP99 PRELIMINARY; PRT; 150 AA.
AC Q9VP99;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN CG13257 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman M.R., Bouck J., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boutin L., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Delcher A., Deng Z., Davenport L.B., Davies P.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Mays A.D., Dietz S.M.,
RA Folsler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodek A., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kallush F., Karpman C., Kravitz S., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AF003592; AAF53541.1; -
DR FLYBASE: FBgn0032587; CG5953.
DR InterPro: IPR001545; Glyco_hormone_beta.
DR Pfam: PF00057; ldl_recept_a; 1.
DR SMART: SM00192; LDLa; 1.
DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; UNKNOWN_1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 1.
KW Glycoprotein.
SQ SEQUENCE 452 AA; 47875 MW; 0F7ABD53014E3E5C CRC64;

Query Match 33.9%; Score 64.5; DB 5; Length 150;
Best Local Similarity 51.6%; Pred. No. 2.9;
Matches 16; Conservative 0; Mismatches 6; Indels 9; Gaps 1;

Qy 5 TLROCL-----AARAGGGGGGGIRGP 26
Db 43 TLLLCGLISLGMATAAARAGGGGGGGGRAP 73

RESULT 22
Q9AF15 ID Q9AF15 PRELIMINARY; PRT; 165 AA.
AC Q9AF15;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN SINGLE-STRANDED DNA-BINDING PROTEIN.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;

```


OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Steward C.A.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;
 RA none;

RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology";
 RL Science 282:2012-2018(1998).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AL032627; CAB54382.1; -

DR InterPro; IPR001871; bZIP.

DR InterPro; IPR003102; pKID.

DR Pfam; PF001170; bZIP; 1.

DR Pfam; PF021173; pKID; 1.

DR SMART; SM00338; BRIZ; 1.

DR PROSITE; PS00036; BZIP_BASIC; 1.

KW DNA-binding; Nuclear protein.

SQ SEQUENCE 333 AA; 35261 MW; BF02CE6398F6D058 CRC64;

Query Match 33.7%; Score 64; DB 5; Length 333;
 Best Local Similarity 76.9%; Pred. No. 6.9;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

DB 169 GGGGGGGVPGPS 181

|||||||: ||:

RESULT 26

O96755 PRELIMINARY; PRT; 422 AA.

AC O96755;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE INTERMEDIATE FILAMENT PROTEIN EI.

OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).

OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

OC Branchiostoma.

OX NCBI_TaxID=7740;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99019308; PubMed=9804163;

RA Karabinos A., Riemer D., Erber A., Weber K.;

RT "Homologues of vertebrate type I, II and III intermediate filament

(IF) proteins in an invertebrate: the IF multigene family of the

cephalochordate Branchiostoma";

RL FEBS Lett. 437:15-18(1998).

DR EMBL; AJ010294; CAA0968.1; -

DR InterPro; IPR002952; Eggshell.

DR InterPro; IPR001664; IF.

DR InterPro; IPR002957; Keratin_I.

DR InterPro; IPR003489; Ribosomal_S30.

Pfam; PF00038; filament; 1.

DR PRINTS; PR01228; EGGSHLL.

DR PRINTS; PR01248; TYPEKERATIN.

SQ SEQUENCE 422 AA; 44892 MW; 85FE742F07751B24 CRC64;

Query Match 33.7%; Score 64; DB 5; Length 422;

Best Local Similarity 61.9%; Pred. No. 8.7;

Matches 13; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY 15 GGGGGGGGIEG-----PTLR 29

DB 92 GGGGGGGGIGMWTKEKPTMR 112

|||||||: ||:

RESULT 27

O9ASE5

ID Q9ASE5

AC Q9ASE5;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)

DE P0456F08.14 PROTEIN.

GN P0456F08.14.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Eriactoidae; Oryzae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

clone:P0456F08.14";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF002901; BAB39414.1; -

DR InterPro; IPR002937; Amino_oxidase.

DR InterPro; IPR000205; NAD_binding.

DR Pfam; PF01593; Amino_oxidase; 1.

SQ SEQUENCE 529 AA; 55981 MW; 0A5DA55CDD076D24 CRC64;

Query Match 33.7%; Score 64; DB 10; Length 529;

Best Local Similarity 51.7%; Pred. No. 11;

Matches 15; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 6 LKCLARAGGGGGGIEGPTLRQCLAA 34

DB 151 LRAYQAARSAGGGGGGKEEVEDEALLA 179

|||||: |||||: ||:

RESULT 28

O9F7T9

ID O9F7T9

AC O9F7T9;

DT 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE AVERMECTIN POLYKETIDE SYNTHASE (FRAGMENT).

OS Streptomyces avermitilis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=33903;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC31271;

RA Hong Y.-S., Lee J.J.;

RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene

and polyketide synthase gene from Streptomyces avermitilis.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF275943; AAG09812.1; -

DR InterPro; IPR001227; Acyltransf_domain.

DR InterPro; IPR000794; Ketoacyl-synt.

DR InterPro; IPR003880; Phosphopant_attach.

Pfam; PF00698; Acyl_transf; 3.

Pfam; PF02801; ketoacyl-synt_C; 2.

Pfam; PF00109; ketoacyl-synt_C; 2.

PROSITE; PS00075; ACP_DOMAIN; 2.

PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.

FT NON_TER 3626 3626

SQ SEQUENCE 3626 AA; 380557 MW; 6272F5F088C1A8D0 CRC64;

Query Match 33.7%; Score 64; DB 2; Length 3626;

Best Local Similarity 54.5%; Pred. No. 67;

Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IEPTLRQCLARAGGGGGGG 22

Wed Oct 9 10:29:58 2002

Query Match 33.4%; Score 63.5; DB 10; Length 113;
Best Local Similarity 36.1%; Pred. No. 2.8;
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 8 QCLAAAR-----AGGGGGGGGIEGPTLRQCLAAARA 36
DB 71 KCMCTKRCGGGGGGGGGGGGGGDEPPLRARVHRS 106

Search completed: October 9, 2002, 09:03:10
Job time : 13.9826 secs

Db 3312 VPGVVVLRCPDAGAGGGGGGG 3333

RESULT 29

Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE I POLYKETIDE SYNTHASE AVES 1.
GN AVEAL.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RL antihelminthic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514 (1999).
DR EMBL; AB032367; BAA84474.1;
DR InterPro: IPR001227; Acyltransferase domain.
DR InterPro: IPR000794; Ketoacyl-synth.
DR InterPro: IPR003880; Phosphopant. attach.
DR InterPro: IPR000834; Zn-carboxypept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS0075; ACP DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 33.7%; Score 64; DB 2; Length 3972;
Best Local Similarity 54.5%; Pred. No. 73;
Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAARAGGGGGGG 22
DB 3345 VPGVVVLRCPDAGAGGGGGGG 3366

RESULT 30

Q942U6 PRELIMINARY; PRT; 113 AA.
AC Q942U6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE P0506E04.26 PROTEIN.
GN P0506E04.26.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0506E04.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF003272; BAB67948.1;
SQ SEQUENCE 113 AA; 11708 MW; 26D9B2C86935BC0B CRC64;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 16.1874 Seconds
(without alignments) 247.023 Million cell updates/sec

Title: US-09-422-838C-29
 Perfect score: 180
 Sequence: 1 IEGPTLRQALAPAGGGGGGIEGPTLRQALARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20

Qst-processing: Minimum

Cost processing: Minimum Match 0%
Maximum Match 100%

Maximum MACCH 100%
Listing first 45 summaries

atabase :

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT: *

2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT: *

3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT: *

4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT: *

5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT: *

6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT: *

7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT: *

8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT: *

9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT: *

10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT: *

11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT: *

12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT: *

13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT: *

14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT: *

15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT: *

16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT: *

17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT: *

18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT: *

19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT: *

20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT: *

21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT: *

22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match		Length	DB	ID	Description
		Query	Match				
1	180	100.0	36	21	AA817300		TPO-mimetic peptid
2	180	100.0	36	21	AA96522		Linear thrombopoietin
3	172	95.6	36	21	AA817298		TPO-mimetic peptid
4	172	95.6	36	21	AA817299		TPO-mimetic peptid
5	172	95.6	36	21	AA96521		Cyclic or linear t
6	166	92.2	36	21	AA816963		TPO-mimetic peptid
7	166	92.2	36	21	AA817293		TPO-mimetic peptid
8	166	92.2	36	21	AA96525		Thrombopoietin mim
9	166	92.2	41	21	AA96528		Thrombopoietin mim
10	166	92.2	42	21	AA817281		TPO-mimetic peptid
11	166	92.2	42	21	AA817282		TPO-mimetic peptid

Seq ID	Accession	Length	Score	Alignment
12	AA017308	42	92.2	100%
13	AA017308	42	92.2	100%
14	AA017308	42	92.2	100%
15	AA017308	42	92.2	100%
16	AA017308	42	92.2	100%
17	AA017308	42	92.2	100%
18	AA017308	42	92.2	100%
19	AA017308	42	92.2	100%
20	AA017308	42	92.2	100%
21	AA017308	42	92.2	100%
22	AA017308	42	92.2	100%
23	AA017308	42	92.2	100%
24	AA017308	42	92.2	100%
25	AA017308	42	92.2	100%
26	AA017308	42	92.2	100%
27	AA017308	42	92.2	100%
28	AA017308	42	92.2	100%
29	AA017308	42	92.2	100%
30	AA017308	42	92.2	100%
31	AA017308	42	92.2	100%
32	AA017308	42	92.2	100%
33	AA017308	42	92.2	100%
34	AA017308	42	92.2	100%
35	AA017308	42	92.2	100%
36	AA017308	42	92.2	100%
37	AA017308	42	92.2	100%
38	AA017308	42	92.2	100%
39	AA017308	42	92.2	100%
40	AA017308	42	92.2	100%
41	AA017308	42	92.2	100%
42	AA017308	42	92.2	100%
43	AA017308	42	92.2	100%
44	AA017308	42	92.2	100%
45	AA017308	42	92.2	100%

ALIGNMENTS

RESULT 1	
AAB17300	
ID AAB17300 standard; Peptide; 36 AA.	
XX	
XX	
AAB17300;	
XX	
XX	
DT 31-OCT-2000 (first entry)	
XX	
DE TPO-mimetic peptide sequence SEQ ID NO:356.	
XX	
KW Modified peptide; therapeutic agent; fusion	
KW autoimmune disease; cytostatic; antiasthma;	
KW immunosuppressive; EPO; TPO; CTLA4; mimetic;	
KW MMP; inhibitor; erythropoietin; thrombopoietic	
KW cytotoxic T cell lymphocyte antigen 4; tumour	
KW vascular endothelial growth factor; matrix m	
KW asthma; thrombosis; pharmaceutical.	
OS Synthetic.	
XX	
XX	
WO2000024782-A2.	
PX	
XX	
04-MAY-2000.	
PD	
XX	
PF 25-OCT-1999; 99WO-US25044.	
XX	
PPR 23-OCT-1998; 98US-0105371.	
XX	
PPR 22-OCT-1999; 99US-0428082.	
XX	
(AMGE-) AMGEN INC.	
XX	
XX	
PI Feige U, Liu C, Cheetham J, Boone TC;	
DR WPI; 2000-350702/30.	

Wed Oct 9 10:29:59 2002

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX
 PS Example 1; Page 321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitumour, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;
 SQ Query Match 100.0%; Score 180; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 2.2e-14;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQALAAAGGGGGGGGGIEPTLRQALAAARA 36
 |||||
 Db 1 IEPTLRQALAAAGGGGGGGGGIEPTLRQALAAARA 36

RESULT 2
 AAY96522
 ID AAY96522 standard; peptide; 36 AA.

XX AAY96522;
 XX
 DT 04-SEP-2000 (first entry)

XX Linear thrombopoietin mimetic peptide compound 3.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; linear.

XX Synthetic.

XX Key Location/Qualifiers
 FT Modified-site 1 /note= "optionally linked to an Fc molecule"
 FT Peptide 1..14
 FT /label= TMP_1
 FT Peptide 15..22
 FT /label= linker
 FT Peptide 23..36
 FT /label= TMP_2

XX WO200024770-A2.

PN 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX

PI Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 61; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2],
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₁-X_{1,1}, X₂-X_{1,1}, X₂-X_{1,1}, X₂-X_{1,1},
 CC X₂-X_{1,1}, X₂-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, and
 CC X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1},
 CC X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1},
 CC X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1}, X₁-X_{1,1},
 CC or E; X₉ = W, Y or S; X₆ = L, I, V, A, F, M, or K; X_{1,1} = A, I, V,
 CC L, F, S, T, K, H, or E; X_{1,2} = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC T, V, N, Q or G; X_{1,4} = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

XX Query Match 100.0%; Score 180; DB 21; Length 36;

XX Best Local Similarity 100.0%; Pred. No. 2.2e-14;

XX Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQALAAAGGGGGGGGGIEPTLRQALAAARA 36
 |||||
 Db 1 IEPTLRQALAAAGGGGGGGGGIEPTLRQALAAARA 36

RESULT 3
 AAB17298

ID AAB17298 standard; Peptide; 36 AA.

XX AAB17298;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:354.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antitumour; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 1; Page 320; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P*3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 36 AA;

Query Match 95.6%; Score 172; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 1.9e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 IEPTLRQALAAAGGGGGGEGPTLRQALAA 36
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEPTLRQCLAAAGGGGGGEGPTLRQCLAA 36

RESULT 4
 AAB17299
 ID AAB17299 standard; Peptide; 36 AA.
 XX
 AC AAB17299;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO.355.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cycostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 FT Key 1 Location/Qualifiers
 FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT Disulfide-bond 1..14 /label= TMP_1
 FT 9..31
 FT /note= "optional"
 FT Peptide 15..22 /label= linker
 FT Peptide 23..36 /label= TMP_2
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 DR WPI; 2000-350702/30.
 XX
 PT Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 1; Page 320-321; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P*3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 36 AA;

Query Match 95.6%; Score 172; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 1.9e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 IEPTLRQALAAAGGGGGGEGPTLRQALAA 36
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEPTLRQCLAAAGGGGGGEGPTLRQCLAA 36

RESULT 5
 AAY96521
 ID AAY96521 standard; peptide; 36 AA.
 XX
 AC AAY96521;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Cyclic or linear thrombopoietin mimetic peptide compound 2.
 XX
 KW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
 XX
 OS Synthetic.
 XX
 FT Key 1 Location/Qualifiers
 FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT Disulfide-bond 1..14 /label= TMP_1
 FT 9..31
 FT /note= "optional"
 FT Peptide 15..22 /label= linker
 FT Peptide 23..36 /label= TMP_2
 XX
 PN WO200024770-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX
 PR 23-OCT-1998; 98US-0105348.
 XX
 PA (AMGE-) AMGEN INC.

Wed Oct 9 10:29:59 2002

LIU C, Feige U, Cheetham J;
WPI; 2000-365108/31.

Thrombopoietic peptides which activate mpl receptors and increase the production of platelets or platelet precursors, useful for treatment of diseases which involve thrombocytopenia

Claim 16; Page 61; 91pp; English.

A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2], is new. TMP-1 and TMP-2 are amino acid sequences varying from at least 10 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂, X₂-X₁-L₃, X₂-X₁-L₄, X₁-X₁-L₀, X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and X₁-X₁-L₄. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S; X₆ = L, I, V, A, F, M, or K; X₇ = R or Q; X₈ = Q, N, or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, T, V, N, Q or G; X₁₂ = A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

Sequence 36 AA;
Query Match 95.6%; Score 172; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 1.9e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQALAAAGAGGGGGGIEGPTLRQALAARA 36
||||| ||||||| ||||||| ||||||| |||||
Db 1 IEPTLRQALAAAGAGGGGGGIEGPTLRQALAARA 36

RESULT 6
AAB16963
ID AAB16963 standard; Protein: 36 AA.
AC AAB16963;
XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide TMP-TMP SEQ ID NO:14.
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.
XX Synthetic.
XX WO200024782-A2.
XX 04-MAY-2000.
XX 25-OCT-1999; 99WO-US25044.
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX (AMGE-) AMGEN INC.
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -
Disclosure; Page 190; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from - (L1)c-P1, - (L1)c-P1-(L2)d-P2, - (L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present inventions are used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAB69443 to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Sequence 36 AA;
Query Match 92.2%; Score 166; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 9.2e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQALAAAGAGGGGGGIEGPTLRQALAARA 36
||||| ||||||| ||||||| ||||||| |||||
Db 1 IEPTLRQALAAAGAGGGGGGIEGPTLRQALAARA 36

RESULT 7
AAB17293
ID AAB17293 standard; Peptide: 36 AA.
XX AAB17293;
XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:349.
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.
XX Synthetic.
XX WO200024782-A2.
XX 04-MAY-2000.
XX 25-OCT-1999; 99WO-US25044.
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX (AMGE-) AMGEN INC.
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and

(PRICE) AMGEN INC.

23-OCT-1998; 98US-0105348.

23-OCT-1998; 98US-0105348.

us-09-422-838c-29.rag

Wed Oct 9 10:29:59 2002

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases.

XX Disclosure; Page 313; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present inventions can

CC be used for producing pharmaceutical compositions. The compositions can

CC be used for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX Sequence 42 AA;

XX Query Match 92.2%; Score 166; DB 21; Length 42;

XX Best Local Similarity 94.4%; Pred. No. 1.le-12;

XX Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLQALAAAGGGGGGGGGGGTPTLQALAAARA 36

DB 7 IEGPTLQALAAAGGGGGGGGGGGTPTLQALAAARA 42

RESULT 11

AAAB17281

ID AAB17282 standard; Peptide: 42 AA.

XX AAB17282;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:338.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

PA (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the

XX production of platelets or platelet precursors, useful for treatment of

XX diseases which involve thrombocytopenia

XX Claim 16; Page 65; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin

XX mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-nmp-2],

CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least

CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,

CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and

CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;

CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N,

CC or E; X-9 = W, Y or F; X-10 = L, I, V, A, F, W, or K; X-11 = A, I, V,

CC L, F, S, T, K, H, or E; X-12 = A, I, V, L, F, T, R, E, or G; X-13 = R, K,

CC T, V, N, Q or G; X-14 = A, I, V, L, F, T, R, E, or G; L-1 = linker

CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

CC activate the c-Mpl receptor which mediates the activity of endogenous

CC thrombopoietin. The TMs are useful for increasing the production of

CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which

CC is useful for treatment of diseases which involve thrombocytopenia, e.g.

CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency

CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 41 AA;

XX Query Match 92.2%; Score 166; DB 21; Length 41;

XX Best Local Similarity 94.4%; Pred. No. 1.le-12;

XX Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLQALAAAGGGGGGGGGGGTPTLQALAAARA 36

DB 6 IEGPTLQALAAAGGGGGGGGGGGTPTLQALAAARA 41

RESULT 10

AAAB17281

ID AAB17281 standard; Peptide: 42 AA.

XX AAB17281;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:337.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Disclosure; Page 313; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AA616955 to AA618003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 42 AA;

Query Match 92.2%; Score 166; DB 21; Length 42;
 Best Local Similarity 94.4%; Pred. No. 1.le-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEPTLRQALAAARAGGGGGGIEGPTLRQALAARA 36
 Db 1 IEPTLRQALAAARAGGGGGGIEGPTLRQALAARA 36

RESULT 12
 AA617308
 ID AA617308 standard; Peptide; 42 AA.
 AC AA617308;
 XX
 DT 31-OCT-2000 (first entry)
 DE Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 PT WPI; 2000-350702/30.
 DR
 XX

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 2; Page 327; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AA616955 to AA618003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 42 AA;

Query Match 92.2%; Score 166; DB 21; Length 42;
 Best Local Similarity 94.4%; Pred. No. 1.le-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 IEPTLRQALAAARAGGGGGGIEGPTLRQALAARA 36
 Db 7 IEPTLRQALAAARAGGGGGGIEGPTLRQALAARA 42

RESULT 13
 AA696530
 ID AA696530 standard; Protein; 42 AA.
 AC AA696530;
 XX
 DT 04-SEP-2000 (first entry)
 DE Thrombopoietin mimetic peptide.
 XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.
 OS Synthetic.
 XX
 PN WO200024770-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX
 PR 23-OCT-1998; 98US-0105348.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Liu C, Feige U, Cheetham J;
 PT WPI; 2000-365108/31.
 DR N-PSDB; AAA29225.
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 PS Example 2A; Page 48; 91pp; English.
 XX

Wed Oct 9 10:29:59 2002

Overlapping oligonucleotides were used to construct a synthetic gene encoding a thrombopoietin mimetic peptide (TMP), which was then fused in-frame to the Fc region of the human IgG1 chain (see AY96529). A compound which binds to an mpl receptor comprising a TMP dimer joined by a linker [TMP-1-(L1)-nTMP-2], is new. TMP-1 and TMP-2 are amino acid sequences varying from at least 10 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂, X₂-X₁-L₃, X₂-X₁-L₄, X₁-X₁-L₀, X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and X₁-X₁-L₄. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N, or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, L, E, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or G; X₁₃ = R, K, T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

Sequence 42 AA; Query Match 92.2%; Score 166; DB 21; Length 42; Best Local Similarity 94.4%; Pred. NO. 1.1e-12; Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQALAAARAGGGGGGIEGPTLRQALAAARA 36
 DB 7 IEGPTLRQALAAARAGGGGGGIEGPTLRQALAAARA 42

RESULT 14
 ID AAB17311 standard; Peptide; 60 AA.

XX AAB17311;
 AC AAB17311;
 DT 31-OCT-2000 (first entry)

XX Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.
 DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.
 OS Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 331; 608pp; English.

XX The present invention describes composition of matter (1) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (1) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 60 AA; Query Match 92.2%; Score 166; DB 21; Length 60;

XX Best Local Similarity 94.4%; Pred. NO. 1.5e-12; Indels 0; Gaps 0;
 XX Matches 34; Conservative 0; Mismatches 2;

QY 1 IEGPTLRQALAAARAGGGGGGIEGPTLRQALAAARA 36
 DB 2 IEGPTLRQALAAARAGGGGGGIEGPTLRQALAAARA 37

RESULT 15
 ID AAB16960 standard; Protein; 269 AA.

XX AAB16960;
 AC AAB16960;

XX 31-OCT-2000 (first entry)

XX TMP-TMP-Fc protein sequence SEQ ID NO:10.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.
 OS Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX N-PSDB; AAA69446.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 185-186; 608pp; English.

pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antitumoric, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences of the present invention.

XX	Sequence	268 AA:
SQ	Query Match	90.0% Score 162; DB 21; Length 268;
	Best Local Similarity	94.3%; Prad. No. 1.9e-11;
	Matches 33; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Oy	1 IEPTTLRQLAARAGGGGGGIEGPTLRQALAAAR 35	
bB	234 IEPTTLRQNLAAARAGGGGGGIEGPTLRQMLAAR 268	

RESULT 18
AAB17301
12 AAB17301 standard: Peptide: 36 AA.

XX	AAE17301;
XX	
XX	31-OCT-2000 (first entry)
DT	
XX	
XX	TPO-mimetic peptide sequence SEQ ID NO.357.
DE	
XX	
XX	Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW	autoimmune disease; cytostatic; antitastmatic; thrombolytic; VEGF;
KW	immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW	MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW	vascular endothelial growth factor; matrix metalloproteinase;
KW	vascular endothelias; pharmaceutical.

XX	
OS	Synthetic.
XX	
XX	
PN	WO200024782-A2.
XX	
XX	
PD	04-MAY-2000.
XX	
EF	25-OCT-1999; 99WO-US25044.
XX	
PR	23-OCT-1998; 98US-0105371.
PR	22-OCT-1999; 99US-0428082.
XX	
DA	{AMGE-} AMGEN INC.

XX	Feige U, Liu C, Cheetham J, Boone TC; .
PI	
XX	
XX	WPI; 2000-350702/30.
DR	
XX	
XX	Novel composition of matter comprising an Fc domain and
PT	pharmacologically active peptides, useful for treating cancer and
PT	autoimmune diseases -
PT	
XX	Patent No. 2000-321, 608pp; English.

XX The present invention describes composition of matter (I) comprising
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-f1-(X2)b, where: f1 = an Fc domain; X1 and X2 = are each
CC independently selected from - (L1)c-p1, - (L1)c-p2, (L2)d-p2,
CC - (L1)c-p1, - (L1)c-p2, (L2)d-p2, (L2)e-p3, or - (L1)c-p1, (L2)d-p2, (L2)e-p3, (L4)f-p4
CC where p1, p2, p3, and p4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, e, and f = are each independently

0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB116955 to AAB18003 represent nucleotide and amino acid sequences of the present invention.

XX	Sequence	36 AA;	87.8%;	Score 158;	DB 21;	Length 36;
SQ						
	Query Match		91.7%;	Pred. No. 7.8e-12;		
	Best Local Similarity					
	Matches 33;	Conservative	0;	Mismatches	3;	Indels
						Gaps
						0;

QY 1 IEGPTLRQALAARAGGGGGGIEGPTLRQALAARA 36

Dh 1 IEGPTLRQALAARAGGGGGGIEGPTLRQALAARA 36

RESULT 19
AAV96523
AAV96523 standard. peptide: 36 AA.

XX	
AC	AAY96523;
XX	
DT	04-SEP-2000 (first entry)
XX	

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW cyclical; linear.

XX	Synthetic.
OS	
XX	
FH	Key
FT	Modified-site
FT	Location/Qualifiers
FT	1 /note= "optionally linked to an Fc molecule"
FT	Peptide
FT	1..14
FT	/label= TMP_1
FT	Peptide
FT	15..22
FT	/label= linker
FT	18
FT	Modified-site
FT	19
FT	20..36
FT	/label= TMP_2
FT	Peptide
FT	23..36
FT	/label= optionally modified by bromoacetyl or PEG"

XX
DN
W0200024770-A2:XX
PD
04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.
PF

23-OCT-1998; 98US-0105348.

PA (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

DR WPI; 2000-365108/31.

Thrombopoietic peptides which activate megakaryocyte precursors, useful for treatment of platelet production of platelets or platelet precursors, useful for treatment of thrombocytopenia

XX
16. Page 62: 91pp: English.

xx
cc A compound which binds to an mpl receptor comprising a thrombopoietin
cc mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-nTMP_2],
cc is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

The present invention describes composition of matter (I) comprising an FC domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)-F1-(X2)_b, where: F1 = an FC domain; X1 and X2 are each independently selected from -(L1)c-p1, -(L1)c-p1-(L2)d-p2, -(L1)c-p1-(L2)d-p2-(L3)e-p*3, or -(L1)c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4 where P1, P2, P3, and P4 = are each independently sequences of

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)-A-(X2)_b, where: P1 = an Fc domain; X1 and X2 = are each independently selected from: -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)b-P2-(L3)e-P3, -(L1)c-P1-(L2)d-P2-(L3)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 37 AA;

Query Match 86.4%; Score 155.5; DB 21; Length 37;
 Best Local Similarity 91.9%; Pred. No. 1.6e-11;
 Matches 34; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 QY 1 IEPTLRQALAAARA--GGGGGGGIEGPTLRQALAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEPTLRQALAAARA--GGGGGGGIEGPTLRQALAAARA 37

RESULT 24
 AAB17295
 ID AAB17295 standard; Peptide; 38 AA.
 XX
 AC AAB17295;
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:351.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.

WO200024782-A2.
 XX
 PN 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;
 PI WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 38 AA;

Query Match 86.1%; Score 155; DB 21; Length 38;
 Best Local Similarity 89.5%; Pred. No. 1.8e-11;
 Matches 34; Conservative 0; Mismatches 2; Indels 2; Gaps 1;
 QY 1 IEPTLRQALAAARA--GGGGGGGIEGPTLRQALAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEPTLRQALAAARA--GGGGGGGIEGPTLRQALAAARA 38

RESULT 25
 AAB17304
 ID AAB17304 standard; Peptide; 39 AA.
 XX
 AC AAB17304;
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:360.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.

WO200024782-A2.
 XX
 PN 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;
 PI WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX Example 1; Page 323; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 36 AA;

Query Match 85.6%; Score 154; DB 21; Length 36;
 Best Local Similarity 88.9%; Pred. No. 2.3e-11;
 Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 IEPTLRQALAAAGGGGGGIEGPTLRQALAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEPTLRQALAAAGGGGGGIEGPTLRQALAAARA 36

RESULT 28

AAY96526
 ID AAY96526 standard; peptide; 36 AA.
 AC AAY96526;
 XX

04-SEP-2000 (first entry)

Thrombopoietin mimetic peptide compound 7.

Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 immunosuppressive; anti-inflammatory; linker.
 XX

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT Peptide 1..14
 FT Peptide 15..18
 FT Peptide /label= linker
 FT Peptide 19..32
 FT Peptide /label= TMP_2

WO200024770-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24834.

23-OCT-1998; 98US-0105348.

(AMGE-) AMGEN INC.

Liu C, Feige U, Cheatham J;

WPI; 2000-365108/31.

Thrombopoietic peptides which activate mpl receptors and increase the
 production of platelets or platelet precursors, useful for treatment of
 diseases which involve thrombocytopenia

Claim 16; Page 62; 91pp; English.

A compound which binds to an mpl receptor comprising a thrombopoietin
 mimetic peptide (TMP) dimer joined by a linker [TMP₁-(L₁)-TMP₂],
 is new. TMP₁ and TMP₂ are amino acid sequences varying from at least
 10 to 14 residues in length comprising X₁-X_{1,0}, X₂-X_{1,1}, X₂-X_{1,2},
 X₂-X_{1,3}, X₂-X_{1,4}, X₁-X_{1,0}, X₁-X_{1,1}, X₁-X_{1,2}, X₁-X_{1,3}, and
 X₁-X_{1,4}. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 X₄ = P; X₅ = T or F; X₆ = L, I, V, A or E; X₇ = R or K; X₈ = Q, N,
 or E; X₉ = W, Y or S; X₁₀ = L, I, V, A, F, M, or K; X_{1,1} = A, I, V,
 L, F, S, T, K, H, or E; X_{1,2} = A, I, V, L, F, T, R, E, or G; X_{1,3} = R, K,
 T, V, N, Q or G; X_{1,4} = A, I, V, L, F, T, R, E, or G; L₁ = linker
 comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 SQ Sequence 36 AA;

Query Match 85.6%; Score 154; DB 21; Length 36;
 Best Local Similarity 88.9%; Pred. No. 2.3e-11;
 Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 IEPTLRQALAAAGGGGGGIEGPTLRQALAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEPTLRQALAAAGGGGGGIEGPTLRQALAAARA 36

RESULT 29

AAB17296
 ID AAB17296 standard; Peptide; 42 AA.
 XX
 AC AAB17296;
 XX

31-OCT-2000 (first entry)

TPO-mimetic peptide sequence SEQ ID NO:352.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 immunosuppressive; EPO; TPO; CTAA4; mimetic; IL-1; TNF; antagonist;
 MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 vascular endothelial growth factor; matrix metalloproteinase;
 asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.

WO200024782-A2.
 04-MAY-2000.

25-OCT-1999; 99WO-US25044.

23-OCT-1998; 98US-0105371.

22-OCT-1999; 99US-0428082.

(AMGE-) AMGEN INC.

Feige U, Liu C, Cheatham J, Boone TC;

WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and
 pharmacologically active peptides, useful for treating cancer and
 autoimmune diseases

Example 1; Page 319; 608pp; English.

The present invention describes composition of matter (I) comprising an
 Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 where P1, P2, P3, and P4 = are each independently sequences of
 pharmacologically active peptides; L1, L2, L3, and L4 = are each
 independently linkers; and a, b, c, d, e, and f = are each
 0 or 1, provided that at least 1 of a and b is 1. The composition can
 have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 activities. DNAs, vectors and host cells from the present invention can
 be used for producing pharmaceutical compositions. The compositions are
 useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 The use of an Fc domain (rather than a Fab domain) can provide a longer

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 ; Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-29
Perfect score: 180
Sequence: 1 IEGPTLRQALAAAGGGGGGIEPTLRQALAAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	65	36.1	738	3	US-08-864-038A-3	Sequence 3, Appl
2	64	35.6	440	3	US-09-100-664A-2	Sequence 2, Appl
3	64	35.6	440	3	US-09-100-664A-3	Sequence 3, Appl
4	64	35.6	440	3	US-09-100-664A-4	Sequence 4, Appl
5	63	35.0	584	2	US-08-987-466-4	Sequence 4, Appl
6	63	35.0	584	2	US-09-240-359-4	Sequence 4, Appl
7	62	34.4	360	2	US-08-319-866-2	Sequence 2, Appl
8	62	34.4	371	2	US-08-442-809A-76	Sequence 76, Appl
9	61	33.9	445	2	US-08-985-090-2	Sequence 2, Appl
10	61	33.9	445	3	US-09-165-543-2	Sequence 2, Appl
11	61	33.9	445	3	US-09-167-354-7	Sequence 7, Appl
12	60.5	33.6	26	1	US-07-776-272-16	Sequence 16, Appl
13	60	33.3	645	4	US-08-791-115B-6	Sequence 6, Appl
14	59.5	33.1	537	1	US-08-472-028A-2	Sequence 2, Appl
15	59.5	33.1	537	2	US-08-808-931-2	Sequence 2, Appl
16	59.5	33.1	537	3	US-08-808-323-2	Sequence 2, Appl
17	59.5	33.1	537	3	US-09-050-603A-2	Sequence 2, Appl
18	59.5	33.1	537	3	US-09-102-420B-2	Sequence 2, Appl
19	59.5	33.1	537	3	US-09-071-296-2	Sequence 2, Appl
20	59.5	33.1	537	4	US-09-196-268-2	Sequence 2, Appl
21	59.5	33.1	537	4	US-09-015-683-2	Sequence 2, Appl
22	59.5	33.1	537	4	US-09-191-998-2	Sequence 2, Appl
23	59.5	33.1	537	4	US-09-497-698-2	Sequence 2, Appl
24	59.5	33.1	969	2	US-08-284-941-2	Sequence 2, Appl
25	59.5	33.1	969	2	US-08-447-642-2	Sequence 2, Appl
26	59.5	33.1	969	4	US-09-236-503-2	Sequence 2, Appl
27	59.5	33.1	969	5	PCT-US93-02147A-2	Sequence 2, Appl

28 59 32.8 14 2 US-08-764-640-13 Sequence 13, Appl
29 59 32.8 14 2 US-08-764-640-193 Sequence 193, Appl
30 59 32.8 14 3 US-08-973-225-13 Sequence 13, Appl
31 59 32.8 14 3 US-08-973-225-193 Sequence 193, Appl
32 59 32.8 14 3 US-09-244-298A-13 Sequence 13, Appl
33 59 32.8 14 3 US-09-244-298A-193 Sequence 193, Appl
34 59 32.8 14 4 US-09-516-704-13 Sequence 13, Appl
35 59 32.8 14 4 US-09-516-704-193 Sequence 193, Appl
36 59 32.8 15 2 US-08-764-640-17 Sequence 17, Appl
37 59 32.8 15 2 US-08-764-640-185 Sequence 185, Appl
38 59 32.8 15 3 US-08-973-225-17 Sequence 17, Appl
39 59 32.8 15 3 US-08-973-225-185 Sequence 185, Appl
40 59 32.8 15 3 US-09-244-298A-17 Sequence 17, Appl
41 59 32.8 15 3 US-09-244-298A-185 Sequence 185, Appl
42 59 32.8 15 4 US-09-516-704-17 Sequence 17, Appl
43 59 32.8 15 4 US-09-516-704-185 Sequence 185, Appl
44 59 32.8 16 2 US-08-764-640-18 Sequence 18, Appl
45 59 32.8 16 2 US-08-764-640-194 Sequence 194, Appl

ALIGNMENTS

RESULT 1
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; CITY: Isshinden
; STATE: Tsu-city
; COUNTRY: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)

us-09-422-838c-29.ra1

Wed Oct 9 10:30:00 2002

```

us-08-864-038A-3
Query Match      36.1%; Score 65; DB 3; Length 738;
Best Local Similarity 58.6%; Pred. No. 3.5;
Matches 17; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 LRQALAAAGGGGGGIEGPTLRQALAA 34
| | | | | | | | | | | | | | | | | |
DB 459 LAAALAAAGAGGGGLGGGGGALAAALAA 487

RESULT 2
US-09-100-664A-2
; Sequence 2, Application US/09100664A
; Patent No. 6057129
; GENERAL INFORMATION:
; APPLICANT: YOUNG, MICHAEL W.
; APPLICANT: KLOSS, BRIAN
; APPLICANT: BLAU, JUSTIN
; APPLICANT: PRICE, JEFFREY
; TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,664A
; FILING DATE: 19-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 440 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-100-664A-3

Query Match      35.6%; Score 64; DB 3; Length 440;
Best Local Similarity 55.0%; Pred. No. 2.7;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 PTLRQALAAAGGGGGGGGGI 23
| | | | | | | | | | | | | |
DB 403 PERRPSIRMQGGGGGGV 422

RESULT 4
US-09-100-664A-4
; Sequence 4, Application US/09100664A
; Patent No. 6057129
; GENERAL INFORMATION:
; APPLICANT: YOUNG, MICHAEL W.
; APPLICANT: KLOSS, BRIAN
; APPLICANT: BLAU, JUSTIN
; APPLICANT: PRICE, JEFFREY
; TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```


us-09-422-838c-29.ra1

Wed Oct 9 10:30:00 2002

```

;
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Yin, Jerry C.
; APPLICANT: Reguski, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/319,866
; FILING DATE: 7-OCT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-319-866-2

Query Match 34.4%; Score 62; DB 2; Length 360;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 ROALAAAGGGGGGGG 22
   :||| |||||
Db 27 QQAATAVGGGGGGG 42

RESULT 8
US-08-442-809A-76
; Sequence 76, Application US/08442809A
; Patent No. 5976873
; GENERAL INFORMATION:
; APPLICANT: Bohinski, Robert J.,
; APPLICANT: Whitsett, Jeffrey A.
; TITLE OF INVENTION: Nucleic Acid Sequences
; TITLE OF INVENTION: Controlling Lung Cell -
; TITLE OF INVENTION: Specific Gene Expression
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS

;
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Yin, Jerry C.
; APPLICANT: Reguski, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/319,866
; FILING DATE: 7-OCT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-319-866-2

Query Match 34.4%; Score 62; DB 2; Length 371;
Best Local Similarity 52.0%; Pred. No. 3.9;
Matches 13; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 7 ROALAAAGGGGGGGG 31
   :||| :|||||
Db 226 QQLQDSDSGGGGGGTCFQQA 250

RESULT 9
US-08-985-090-2
; Sequence 2, Application US/08985090
; Patent No. 5885893
; GENERAL INFORMATION:
; APPLICANT: Andrew D.J. Goodearl
; TITLE OF INVENTION: MUSCARINIC RECEPTORS AND USES THEREFOR
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/985,090
; FILING DATE:
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jean M. Silveri
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: MNI-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,809A
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,356
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Olsteind, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 271010-360
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: human thyroid transcription factor-1
; US-08-442-809A-76
```

```
; MOLECULE TYPE: protein
US-08-985-090-2

Query Match      33.9%; Score 61; DB 2; Length 445;
Best Local Similarity 42.9%; Pred. No. 6;
Matches 12; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 8 QALAAAGGGGGGGIEGPTLRQALAAAR 35
   :| | | | | | | | | | | | | | | |
Db 285 EAGATLGGGGGGVSPSTSSGSSSR 312

RESULT 10
US-09-165-543-2
; Sequence 2, Application US/09165543
; Patent No. 6093545
; GENERAL INFORMATION:
; APPLICANT: Andrew D.J. Goodearl and Sandra Glucksman
; TITLE OF INVENTION: Muscarinic Receptors and Uses Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/165,543
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,780
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth A. Hanley
; REGISTRATION NUMBER: 33,505
; REFERENCE/DOCKET NUMBER: MNI-032CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-165-543-2

Query Match      33.9%; Score 61; DB 3; Length 445;
Best Local Similarity 42.9%; Pred. No. 6;
Matches 12; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 8 QALAAAGGGGGGGIEGPTLRQALAAAR 35
   :| | | | | | | | | | | | | | | |
Db 285 EAGATLGGGGGGVSPSTSSGSSSR 312

RESULT 11
US-09-167-354-7
; Sequence 7, Application US/09167354A
; Patent No. 6136559
; GENERAL INFORMATION:
; APPLICANT: Lovenberg, Timothy
; APPLICANT: Erlander, Mark
; APPLICANT: Pyati, Jayashree
; APPLICANT: Huvar, Arne
; TITLE OF INVENTION: DNA ENCODING A HUMAN HISTAMINE RECEPTOR OF THE H3
```

```
; TITLE OF INVENTION: SUBTYPE
; FILE REFERENCE: JMW
; CURRENT APPLICATION NUMBER: US/09/167,354A
; CURRENT FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:PEPTIDE
US-09-167-354-7

Query Match      33.9%; Score 61; DB 4; Length 445;
Best Local Similarity 42.9%; Pred. No. 6;
Matches 12; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 8 QALAAAGGGGGGGIEGPTLRQALAAAR 35
   :| | | | | | | | | | | | | | | |
Db 285 EAGATLGGGGGGVSPSTSSGSSSR 312

RESULT 12
US-07-776-272-16
; Sequence 16, Application US/07776272
; Patent No. 5612454
; GENERAL INFORMATION:
; APPLICANT: Kaminuma, Toshihiko
; APPLICANT: Iida, Toshii
; APPLICANT: Tajima, Masahiro
; TITLE OF INVENTION: Process for Purification of Polypeptide
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player
; STREET: 1233 20th St. N.W. P.O. Box 18218
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20036-8218
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/776,272
; FILING DATE: 19911129
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P-450-23167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-0400
; TELEFAX: 202-887-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Bovine
US-07-776-272-16

Query Match      33.6%; Score 60.5; DB 1; Length 26;
Best Local Similarity 50.0%; Pred. No. 0.39;
Matches 13; Conservative 2; Mismatches 8; Indels 3; Gaps 1;

QY 1 IEGPTLRQALAAAGGGGGIEGP 26
```


NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1847
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-931-2

Query Match 33.1%; Score 59.5; DB 2; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;
QY 3 GPTLRQALAAARAGGGG-----GGGIEGPTLRQALAAAR 35
|||: ::: |||| | : |||| :
Db 39 GPTVG---SSKIEGGGTTTDCVIVGGISGLCIAQALATK 78

RESULT 16
US-08-808-323-2
Sequence 2, Application US/08808323
Patent No. 6018105
GENERAL INFORMATION:
APPLICANT: Johnson, Marie
APPLICANT: Volrath, Sandra
APPLICANT: Ward, Eric
TITLE OF INVENTION: Promoters from Plant
TITLE OF INVENTION: Protoporphyrinogen Oxidase Genes
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6018105artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591-9005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,323
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/012,705
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,612
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/020,003
FILING DATE: 21-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1846
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-323-2

Query Match 33.1%; Score 59.5; DB 3; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;
QY 3 GPTLRQALAAARAGGGG-----GGGIEGPTLRQALAAAR 35
|||: ::: |||| | : |||| :
Db 39 GPTVG---SSKIEGGGTTTDCVIVGGISGLCIAQALATK 78

RESULT 17
US-09-050-603A-2
Sequence 2, Application US/09050603A
Patent No. 6023012
GENERAL INFORMATION:
APPLICANT: Volrath, Sandra
APPLICANT: Johnson, Marie
APPLICANT: Potter, Sharon
APPLICANT: Ward, Eric
APPLICANT: Helfetz, Peter
TITLE OF INVENTION: DNA Molecules Encoding Plant
TITLE OF INVENTION: Protoporphyrinogen Oxidase
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6023012artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,603A
FILING DATE: 30-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/808,931
FILING DATE: 28-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/012,705
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,612
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/020,003
FILING DATE: 21-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1847
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-050-603A-2

Query Match 33.1%; Score 59.5; DB 3; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;
QY 3 GPTLRQALAAARAGGGG-----GGGIEGPTLRQALAAAR 35
|||: ::: |||| | : |||| :
Db 39 GPTVG---SSKIEGGGTTTDCVIVGGISGLCIAQALATK 78

RESULT 18
US-09-102-420B-2
; Sequence 2, Application US/09102420B
; Patent No. 6084155
; GENERAL INFORMATION:
; APPLICANT: Volrath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Ward, Eric
; APPLICANT: Helfetz, Peter
; TITLE OF INVENTION: HERBICIDE-TOLERANT PROTOPORPHYRINOGEN
; TITLE OF INVENTION: OXIDASE ("PROTOX")
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6084155artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,420B
; FILING DATE: 22-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/059,164
; FILING DATE: 13-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/050,603
; FILING DATE: 30-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/126,430
; FILING DATE: 11-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/808,931
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/012,705
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,612
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,003
; FILING DATE: 21-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/472,028
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1847/CIP3
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-102-420B-2

Query Match 33.1%; Score 59.5; DB 3; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;

QY 3 GPTLRQALAAARAGGGG-----GGGIEGPTLRQALAAAR 35
|||: :: |||| | : ||| :
Db 39 GPTVG---SSKIEGGGTTITDCVIVGGISGLCIAQALATK 78
|||: :: |||| | : ||| :
RESULT 19
US-09-071-296-2
; Sequence 2, Application US/09071296
; Patent No. 6177245
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,296
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-296-2

Query Match 33.1%; Score 59.5; DB 4; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;

QY 3 GPTLRQALAAARAGGGG-----GGGIEGPTLRQALAAAR 35
|||: :: |||| | : ||| :
Db 39 GPTVG---SSKIEGGGTTITDCVIVGGISGLCIAQALATK 78
|||: :: |||| | : ||| :
RESULT 20
US-09-196-268-2
; Sequence 2, Application US/09196268
; Patent No. 6282837
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY

; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,268
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-196-268-2

Query Match 33.1%; Score 59.5; DB 4; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;
QY 3 GPTLRQALAAARAGGGG-----GGGEGPTLRQALAAAR 35
||| : : : ||| : ||| : ||| :
Db 39 GPTVG---SSKIEGGGTTITDCVIVGGISGLCIAQALATK 78

RESULT 21
US-09-015-683-2
; Sequence 2, Application US/09015683
; Patent No. 6288306
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Vollrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,683
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-015-683-2

Query Match 33.1%; Score 59.5; DB 4; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;
QY 3 GPTLRQALAAARAGGGG-----GGGEGPTLRQALAAAR 35
||| : : : ||| : ||| : ||| :
Db 39 GPTVG---SSKIEGGGTTITDCVIVGGISGLCIAQALATK 78

RESULT 22
US-09-191-998-2
; Sequence 2, Application US/09191998
; Patent No. 6307129
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Vollrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,998
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-191-998-2

Query Match 33.1%; Score 59.5; DB 4; Length 537;
Best Local Similarity 37.2%; Pred. No. 11;
Matches 16; Conservative 6; Mismatches 8; Indels 13; Gaps 2;
QY 3 GPTLRQALAAARAGGGG-----GGGEGPTLRQALAAAR 35
||| : : : ||| : ||| : ||| :
Db 39 GPTVG---SSKIEGGGTTITDCVIVGGISGLCIAQALATK 78

RESULT 23
US-09-497-698-2
; Sequence 2, Application US/09497698
; Patent No. 6308458
; GENERAL INFORMATION:

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND

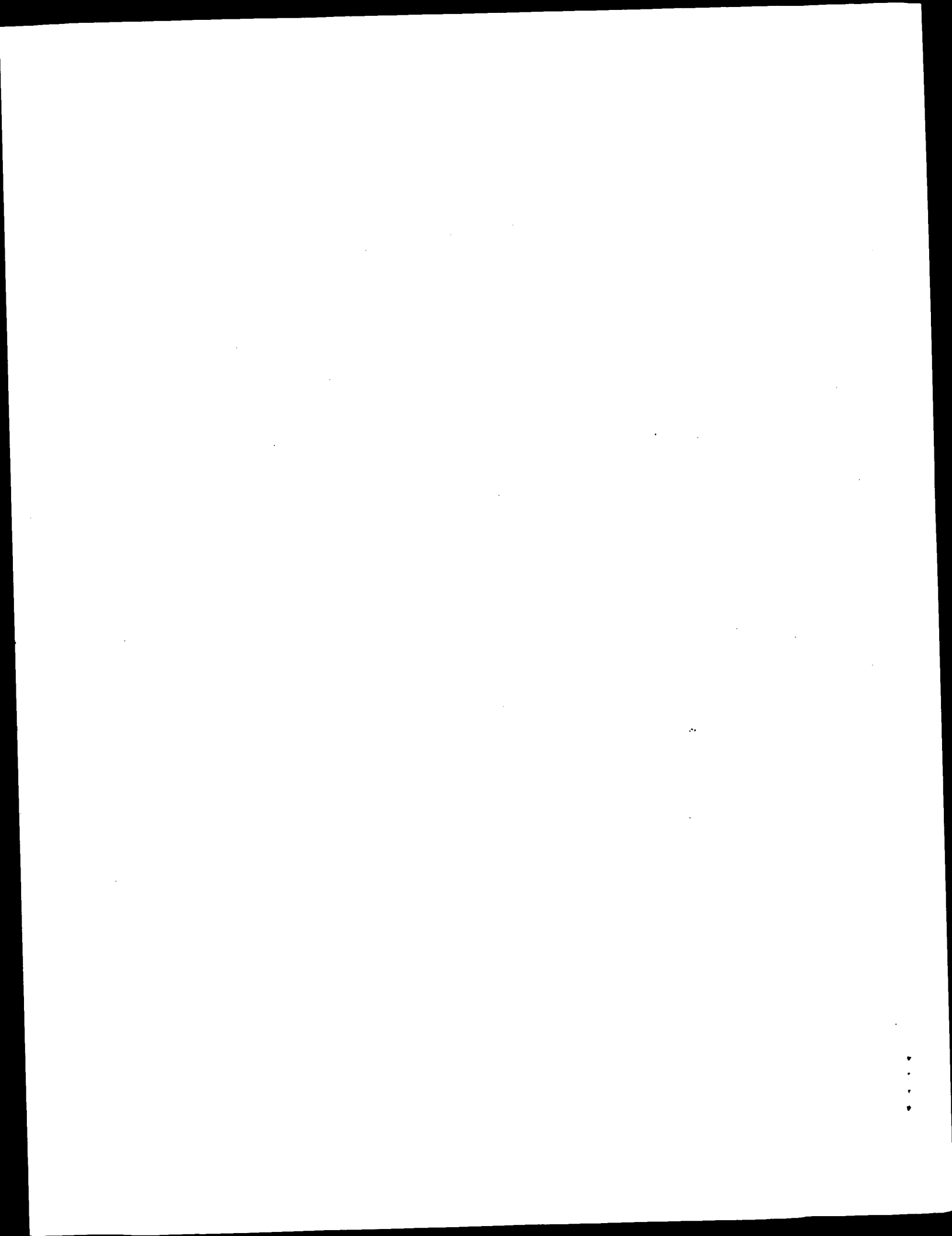
Wed Oct 9 10:30:00 2002

us-09-422-838c-29.rai

Page 13

QY 1 IEGPTLRQALAARA 14
 |||||||
Db 1 IEGPTLRQALAARA 14

Search completed: October 9, 2002, 09:06:33
Job time : 6.98595 secs




```

:923-978/Region: alanine/glycine/proline-rich
:983-1030/Domain: SH3 homology <SH3>
:1034-1168/Region: alanine/glycine/proline-rich
:107/Binding site: ATP (Lys) #status Predicted
:311/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match          37.28;  Score 67;  DB 1;  Length 1168;
Best Local Similarity 50.0%;  Pred. No. 12;
Matches 14;  Conservative 4;  Mismatches 10;  Indels 0;  Gaps 0;

QY      8 QALAAKAGGGGGGGGIEGPTLRQALAAK 35
      | | | | | | | | | | | | | | | |
DB      920 QILGAKGGGGGGGRGGPSPGAVSPR 947

RESULT 5
JQ1094
hypothetical 20.2K protein - tomato ringspot virus
C:Species: tomato ringspot virus
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Oct-1999
C:Accession: JQ1094
R:Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
J. Gen. Virol. 72, 1505-1514, 1991
A:Title: Nucleotide sequence of tomato ringspot virus RNA-2.
A:Reference number: JQ1093; MUID:91311402
A:Accession: JQ1094
A:Status: translation not shown
A:Molecule type: genomic RNA
A:Residues: 1-201 <ROT>
A:Cross-references: GB:D12477; GB:D01129; MID:q222674; PIDN:BAA02044.1; PID:d1002526;
A:Experimental source: strain raspberry

Query Match          36.7%;  Score 66;  DB 2;  Length 201;
Best Local Similarity 51.5%;  Pred. No. 3;2;
Matches 14;  Conservative 4;  Mismatches 10;  Indels 0;  Gaps 0;

```

Query Match	36.7%	Score 66;	DB 2;	Length 201;
Best Local Similarity	61.5%	Pred. No. 3.2;		
			5. Index/s	4. Gaps
				1;

13 RAGGGGGGGIE---GPTLRQALAA 34
13 RAGGGGGGGKEVFKAGRTLKVLKA 38

RESULT 6

T20961
 Hypothetical protein F15B9.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20961
 R:Percy, C.
 Submitted to the EMBL Data Library, August 1996
 A:Reference number: Z19351
 A:Accession: T20961
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-500 <Wtl>
 A:Cross-references: EMBL:Z78013; PIDN:CA801420.1; GSFDB:GN00023; CESP:F15B9.5
 A:Experimental source: Clone F15B9
 C:Genetics:
 A:Gene: CESP:F15B9.5
 A:Map position: 5
 A:Introns: 46/3; 63/3; 125/2; 162/2; 283/3; 391/1; 446/1

Query Match	36.7%	Score 66;	DB 2;	Length 500;
Best Local Similarity	56.5%	pred. No. 7.3;		
Matches 13; Conservative	3;	Mismatches	7;	Indels 0; Gaps 0;

QY 3 GPTLRQALAAARAGGGGGGGGIEG 25
| | : | | ||||| : |
... GATTCCTTCCCGCCGGGGMG 451

RESULT 7
GG87033
probable ATP/GTP-binding protein [imported] - Mycobacterium leprae

```

QY      3 GPTLRQAL-----AARAGGGGGGIEGPTLRLQAARA 36
       ||::| : ||||| |::||:| |
Db     75 GPESKRSIGSPYANNSSNGSGGGGGPGGASITQIAQAA 121

RESULT 3
S$7795
probable deoxyribodipyrimidine photo-lyase (EC 4.1.99.3) - Chlamydomonas reinhardtii
N:Alternate names: DNA photolyase homolog; probable blue light photoreceptor
C:Species: Chlamydomonas reinhardtii
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S$7795; S66368
R:Small, G.D.; Min, B.; Lefebvre, P.A.
plant Mol. Biol. 28, 443-454, 1995
A:title: Characterization of a Chlamydomonas reinhardtii gene encoding a protein of the
A:Reference number: S$7795; MUID:95359403
A:Accession: S$7795
A:Molecule type: DNA
A:Residues: 1-867 <SMW>
A:Cross-references: EMBL:L07561; NID:g945420; PIDN:AAC37438.1; PID:g945421
A:Accession: S66368
A:Molecule type: mRNA
A:Residues: 1-867 <SMW>
A:Cross-references: EMBL:L07561; NID:g945420; PIDN:AAC37438.1; PID:g945421
C:Genetics:
A:Gene: CPH1
A:Introns: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3
A:Exons: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3
A:Protein: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3
A:Enzyme: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3
A:Function: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3
A:Description: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3
A:Keywords: 34/3; 159/2; 210/3; 265/3; 329/3; 406/1; 837/3

```

Query Match 38.1%; Score 68.5; DB 2; Length 867;
Best Local Similarity 50.0%; Pred. No. 6.6;
Matches 15; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

4 PTLRQALAAARAGGGGGGGGIEG-PTLRQAL 32

RESULT 4

MYOSIN heavy chain IC - Acanthamoeba castellanii
 N:Contains: myosin ATPase (EC 3.6.1.32)
 C:Species: Acanthamoeba castellanii
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
 C:Accession: A33891; C34448; A24146
 R:Jung, G.; Korn, E.D.; Hammer III, J.A.
 proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
 A:title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like and non-my
 A:Reference number: A33891; MUID:88016163
 A:Accession: A33891
 A:Molecule type: DNA
 A:Residues: 1-1168 <JUN>
 A:Cross-references: GBJ02974; NID:gl55624; PIDN:AAA27707.1; PID:gl55625
 A:Note: this gene and protein are called MIB in this paper
 R:Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
 J. Biol. Chem. 264, 19340-19348, 1989
 A:title: The localization and sequence of the phosphorylation sites of Acanthamoeba myos
 A:Reference number: A34448; MUID:90037074
 A:Accession: C34448
 A:Molecule type: Protein
 A:Residues: 308-314, 'X', 316-329 <BR2>
 C:Comment: In this protein, the coiled-coil rod-like region found in many myosin heavy c
 he protein is globular and does not self-associate into filaments.
 C:Genetics:
 A:Gene: MIC
 A:Introns: 1/3; 37/3; 60/2; 153/3; 179/3; 208/2; 242/3; 287/3; 321/3; 371/3; 428/3
 C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homolo
 C:Keywords: actin binding; ATP; hydrolyase; nucleotide binding; P-loop; phosphoprotein; t
 F:10-653/Domain: myosin motor domain homology <MMOT>
 F:101-108/Region: nucleotide-binding motif A (P-loop)
 F:543-564/Region: actin binding #status predicted
 F:671-1168/Domain: carboxyl-terminal <CTD>
 F:675-883/Region: basic

A: Experimental source: clone Y41C4A
 B: GenBank
 C: Genetics
 A: Gene: CBSP:Y41C4A.4a
 A: Introns: 24/3; 50/2; 81/3; 159/1; 228/1; 292/3
 C: Superfamily: fos/jun DNA-binding domain homology
 Query Match 35.6%; Score 64; DB 2; Length 331;
 Best Local Similarity 76.9%; Pred. No. 8.1;
 Matches 10; Conservative 1; Indels 0; Gaps 0;
 QY 15 GGGGGGGGIEGPT 27
 LLLLLLLLLL: LL
 DB 167 GGGGGGGVPGPS 179
 RESULT 14
 T26808
 hypothetical protein Y41C4A.4b - *Caenorhabditis elegans*
 C: Species: *Caenorhabditis elegans*
 C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
 C: Accession: T26808
 R: Steward, C. The EMBL Data Library, October 1998
 submitted to the EMBL Data Library, October 1998
 A: Reference number: Z20269
 A: Accession: T26808
 A: Status: preliminary; translated from GB/EMBL/DDBJ
 A: Molecule type: DNA
 A: Residues: 1-333 <WIL>
 A: Cross-references: EMBL:AL032627; PIDN:CAB54382.1; CBSP:Y41C4A.4b
 A: Experimental source: clone Y41C4A
 C: Genetics:
 A: Gene: CBSP:Y41C4A.4b
 A: Introns: 24/3; 50/2; 81/3; 161/1; 230/1; 294/3
 C: Superfamily: fos/jun DNA-binding domain homology
 Query Match 35.6%; Score 64; DB 2; Length 333;
 Best Local Similarity 76.9%; Pred. No. 8.2;
 Matches 10; Conservative 1; Indels 0; Gaps 0;
 QY 15 GGGGGGGGIEGPT 27
 LLLLLLLLLL: LL
 DB 169 GGGGGGGVPGPS 181
 RESULT 15
 C82844
 alanyl-tRNA synthetase XF0124 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C: Species: *Xylella fastidiosa*
 C: Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C: Accession: C82844
 R: anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 408, 151-157, 2000
 A: Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A: Reference number: A92515; MUID:20365717
 A: Note: for a complete list of authors see reference number A59328 below
 A: Accession: C82844
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-916 <SIM>
 A: Cross-references: GB:AE003866; GB:AE003849; NID:99104906; PIDN:AAF62937.1; GSPDB:GN
 A: Experimental source: strain 9a5c
 R: Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A: Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; La
 J.D.; Junqueira, M.L.; Kemper, E.B.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A: Authors: Martins, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Sa, R.G.; Santelli, R.V.; Sawa
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Silva Jr., W.A.; da Sil
 A: Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
 M.; Tshako, M.N.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

1779
ycine-rich RNA-binding protein GRP1 - wheat
Species: Triticum aestivum (common wheat)
Date: 04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 23-Jul-1999
Accession: S71779
Gulltunan, M.J.; Niu, X.
Ant Mol. Biol. 30, 1301-1306, 1996
Title: cDNA encoding a wheat (Triticum aestivum cv. Chinese spring) glycine-rich RNA-b
Reference number: S71779; M01D:96311016
Accession: S71779
Status: preliminary
Molecule type: mRNA
Residues: 1-167 <GUI>
Cross-references: EMBL:U32310; NID:g974604; PIDN:AAA75104.1; PID:g974605
Genetics:
;Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology
;74/Domain: ribonucleoprotein repeat homology <RRM4>
Query Match 35.6%; Score 64; DB 2; Length 167;
Best Local Similarity 61.1%; Pred. No. 4.4; 3; Indels 0; Gaps 0;
Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
5 TLRQALNARAGGGGGG 22
78 TVNEAARRSGGGGGG 95
YY
bb
RESULT 12
S69312
probable membrane protein YLR338w - yeast (Saccharomyces cerevisiae)
;Alternate names: hypothetical protein L8300.13-a
;Species: Saccharomyces cerevisiae
Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 05-Nov-1999
Accession: S69312
Molecule type: DNA
Residues: 1-285 <DUZ>
Cross-references: EMBL:U19028; NID:g609380; PID:g2340034; GSPDB:GN00012; MIPS:YLR338w
R;Du, Z.
submitted to the EMBL Data Library, January 1994
Description: The sequence of S. cerevisiae cosmid 8300.
Reference number: S69312
Accession: S69312
Molecule type: DNA
Residues: 1-285 <DUZ>
Cross-references: EMBL:U19028; NID:g609380; PID:g2340034; GSPDB:GN00012; MIPS:YLR338w
Genetics:
A;Gene: MIPS:YLR338w
A;Map position: 12R
C;Keywords: transmembrane protein
C;142-158/Domain: transmembrane #status predicted <TM1>
F;201-217/Domain: transmembrane #status predicted <TM2>
Query Match 35.6%; Score 64; DB 2; Length 285;
Best Local Similarity 60.0%; Pred. No. 7.1;
Matches 13; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 9 AALARAGGGGGGGIEGPTL 28
DB 235 ALPPNAGGGGGGGAGAPAI 254
RESULT 13
T26807
hypothetical protein Y41C4A.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
Accession: T26807
R;Steward, C.
submitted to the EMBL Data Library, October 1998
Reference number: Z20269
Accession: T26807
Status: preliminary; translated from GB/EMBL/DDBB
Molecule type: DNA

Wed Oct 9 10:30:01 2002

A:Molecule type: DNA
 A:Residues: 137-777 <Q12>
 A:Cross-references: EMBL:X55174
 A:Accession: S65542
 A:Molecule type: DNA
 A:Residues: 'MQAEO', 86-87, 'IG', 90-91, 'QKHYSLKKNRHTLANVR', 94-777 <Q13>
 A:Cross-references: EMBL:X55174
 A:Accession: S65544
 A:Molecule type: DNA
 A:Residues: 'MVCFCFCCCNFRN', 4, 'P', 6, 'S', 94-777 <Q14>
 A:Cross-references: EMBL:X55174
 R:Chen, C.N.; Denome, S.; Davis, R.L.
 Proc. Natl. Acad. Sci. U.S.A. 83, 9313-9317, 1986
 A:Title: Molecular analysis of cDNA clones and the corresponding genomic coding sequence
 A:Reference number: A26651; MUID:87092243
 A:Accession: A26651
 A:Molecule type: DNA
 A:Residues: 416-777 <CHE>
 A:Cross-references: GB:M14982; NID:gl57278; PIDN:AAC34201.1; PID:gl57280
 C:Genetics:
 A:Gene: FlyBase:dnc; dunc
 A:Cross-references: FlyBase:FBgn0000479
 A:Introns: 93/2; 125/3; 152/2; 165/2; 200/2; 262/3; 294/1; 407/3; 496/2; 534/2; 588/3; 7
 C:Superfamily: 3',5'-cyclic-nucleotide phosphodiesterase 1B, calmodulin-dependent; 3',5'
 C:Keywords: alternative splicing; phosphoric diester hydrolase
 F:137-777/Product: cAMP-dependent 3',5'-cyclic-nucleotide phosphodiesterase, splice form
 F:439-667/Domain: 3',5'-cyclic-nucleotide phosphodiesterase homology <CNP>

 Query Match 35.0%; Score 63; DB 2; Length 777;
 Best Local Similarity 72.2%; Pred. No. 22;
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

 Qy 9 ALAARAGGGGGGIEGP 26
 | | | | | | | | | |
 Db 746 ASALRAGGGGGGGMAP 763

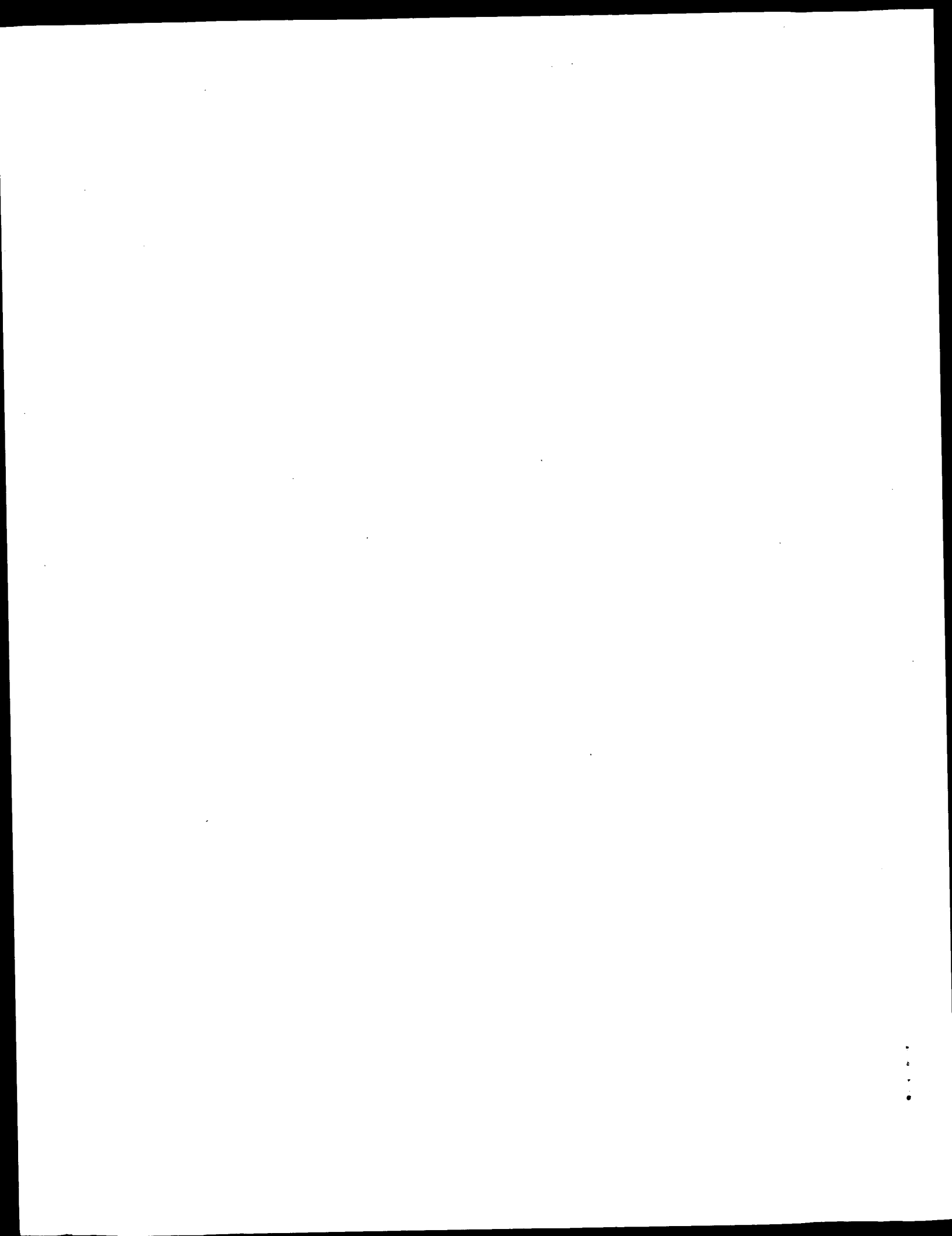
 RESULT 21
 H83533
 alanyl-tRNA synthetase PA0903 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83533
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337
 A:Accession: H83533
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-874 <STO>
 A:Cross-references: GB:AE004524; GB:AE004091; NID:g9946795; PIDN:AAG04292.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: alas; PA0903
 C:Superfamily: alanine--trna ligase

 Query Match 35.0%; Score 63; DB 2; Length 874;
 Best Local Similarity 43.2%; Pred. No. 25;
 Matches 16; Conservative 4; Mismatches 11; Indels 6; Gaps 1;

 Qy 3 GPTLRQALAAAGGGG-----GGGIEGPTLRQALA 33
 | : | | | | | | | | | | | | | | | |
 Db 828 GELMKQAAAVGGGGRPDMAAGGGTDAAKLDEALA 864

 RESULT 22
 IP80
 insulin precursor - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 24-Apr-1984 #sequence_revision 22-Apr-1995 #text_change 16-Jul-1999

C:Accession: A40909; A92080; A91185; A90342; A90341; S48184; S48185; S46258;
 R:D'Agostino, J.; Younes, M.A.; White, J.W.; Besch, P.K.; Field, J.B.; Frazier, M.L.
 Mol. Endocrinol. 1, 327-331, 1987
 A:Title: Cloning and nucleotide sequence analysis of complementary deoxyribonucleic a
 A:Reference number: A40909; MUID:88288209
 A:Accession: A40909
 A:Molecule type: mRNA
 A:Residues: 1-105 <DAA>
 A:Cross-references: GB:M54979; NID:gl63578; PIDN:AAA30722.1; PID:gl63579
 A:Experimental source: fetal pancreas
 R:Nolan, C.; Margoliash, E.; Peterson, J.D.; Steiner, D.F.
 J. Biol. Chem. 246, 2780-2795, 1971
 A:Title: The structure of bovine proinsulin.
 A:Reference number: A92080; MUID:71166442
 A:Accession: A92080
 A:Molecule type: protein
 A:Residues: 25-105 <NOL>
 R:Steiner, D.F.; Cho, S.; Oyer, P.E.; Terris, S.; Peterson, J.D.; Rubenstein, A.H.
 J. Biol. Chem. 246, 1365-1374, 1971
 A:Title: Isolation and characterization of proinsulin C-peptide from bovine pancreas.
 A:Reference number: A92074; MUID:7116409
 A:Accession: A92074
 A:Molecule type: protein
 A:Residues: 57-82 <STE>
 R:Salokangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 183-189, 1971
 A:Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancre
 A:Reference number: A91185; MUID:71257721
 A:Accession: A91185
 A:Molecule type: protein
 A:Residues: 57-82 <SAL>
 R:Sanger, F.; Thompson, E.O.P.
 Biochem. J. 53, 366-374, 1953
 A:Title: The amino-acid sequence in the glycol chain of insulin. 2. The investigation
 A:Reference number: A90342
 A:Accession: A90342
 A:Molecule type: protein
 A:Residues: 85-105 <SAN>
 R:Sanger, F.; Tuppy, H.
 Biochem. J. 49, 481-490, 1951
 A:Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investi
 A:Reference number: A90341
 A:Accession: A90341
 A:Molecule type: protein
 A:Residues: 25-54 <SA2>
 R:Cheng, R.; Kawakishi, S.
 Eur. J. Biochem. 223, 759-764, 1994
 A:Title: Site-specific oxidation of histidine residues in glycosylated insulin mediated b
 A:Reference number: S48184; MUID:94333378
 A:Accession: S48184
 A:Molecule type: protein
 A:Residues: 85-105 <CHE>
 A:Accession: S48185
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 25-30, 'X', 32-42, 'X', 44-54 <CH2>
 R:Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.
 Biochem. J. 60, 541-556, 1955
 A:Title: The disulphide bonds of insulin.
 A:Reference number: A90343
 A:Contents: annotation; amides; disulfides
 R:Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W.
 FEBS Lett. 349, 205-209, 1994
 A:Title: Existence of a molecular ruler in proteasomes suggested by analysis of degra
 A:Reference number: S46258; MUID:94326921
 A:Accession: S46258
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 25-54 <WEN>
 C:Superfamily: insulin
 C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <BCH>



GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-29

Perfect score: 180

Sequence: 1 IEGPTLRQALAAAGGGGGGIEGPTLRQALAAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	68.5	38.1	394	1 FXD3_CHICK	P79772 gallus gall
2	67	37.2	1168	1 MYSC_ACACA	P10569 acanthamoeb
3	66.5	36.9	4499	1 DYHA_CHLRE	Q39610 chlamydomon
4	66	36.7	201	1 YR21_TRSVR	P25245 tomato ring
5	65	36.1	619	1 LAC1_NEUCR	P06811 neurospora
6	65	36.1	619	1 LAC2_NEUCR	P10574 neurospora
7	65	36.1	1178	1 PHYB_SORBI	P93527 sorghum bic
8	64	35.6	440	1 DCO_DROME	Q76324 drosophila
9	63.5	35.3	882	1 SYA_THETH	P74941 thermus aqu
10	63	35.0	323	1 JUND_CHICK	P27921 gallus gall
11	63	35.0	584	1 CNAL_DROME	P12252 drosophila
12	62.5	34.7	105	1 INS_BOVIN	P01317 bos taurus
13	62.5	34.7	105	1 INS_SHEEP	P01318 ovis aries
14	62.5	34.7	168	1 SSB_MYCLE	P46390 mycobacteri
15	62	34.4	112	1 TTF1_CAVPO	P97273 cavia porce
16	62	34.4	371	1 TTF1_CANFA	P43698 canis famil
17	62	34.4	371	1 TTF1_HUMAN	P43698 canis famil
18	62	34.4	476	1 EVX2_HUMAN	Q03828 homo sapien
19	62	34.4	497	1 FXD2_HUMAN	Q60548 homo sapien
20	61.5	34.2	401	1 HB2_HUMAN	P50219 homo sapien
21	61	33.9	445	1 HH3R_HUMAN	Q9Y5N1 homo sapien
22	61	33.9	485	1 FXD1_HUMAN	O16676 homo sapien
23	61	33.9	485	1 ONC2_HUMAN	Q95948 homo sapien
24	61	33.9	517	1 Y967_TREPA	O83933 treponema p
25	61	33.9	753	1 ZIN_HUMAN	Q9N133 homo sapien
26	60	33.3	348	1 SXL_CERCA	Q61374 ceratitidis c
27	60	33.3	421	1 BR3A_MOUSE	P17208 mus musculu
28	60	33.3	569	1 KICJ_MOUSE	P02535 mus musculu
29	60	33.3	796	1 KF3C_RAT	O55165 rattus norv
30	60	33.3	1322	1 SUS_DROME	P22293 drosophila
31	59.5	33.1	301	1 CC02_CAEEL	P17656 caenorhabdi
32	59.5	33.1	391	1 SOX1_MOUSE	P53783 mus musculu
33	59.5	33.1	537	1 PPOC_ARATH	P55826 arabidopsis

34	59.5	33.1	872	1 CIO3_HUMAN	O43525 homo sapien
35	59.5	33.1	969	1 PAC4_HUMAN	P29122 homo sapien
36	59.5	33.1	1122	1 HDA5_HUMAN	Q9UQ16 homo sapien
37	59	32.8	367	1 BET3_MESAU	O09029 mesocricetu
38	59	32.8	369	1 TMAF_AVIS4	P23091 avian muscu
39	59	32.8	419	1 K2C2_XENLA	P04365 xenopus lae
40	59	32.8	448	1 OCT6_HUMAN	O03052 homo sapien
41	59	32.8	449	1 OCT6_MOUSE	P21952 mus musculu
42	59	32.8	451	1 OCT6_RAT	P20267 rattus norv
43	59	32.8	587	1 CH60_BRANA	P35480 brassica na
44	59	32.8	593	1 KICJ_HUMAN	P13645 homo sapien
45	59	32.8	644	1 XYND_CELFI	P54865 cellulomona

ALIGNMENTS

RESULT 1
FXD3_CHICK
ID FXD3_CHICK STANDARD; PRT; 394 AA.
AC P79772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor genesis) (Winged
DE helix protein CWH-3).
GN FOXD3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=97141794; PubMed=8988052;
RA Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;
RT "Aberrant cell growth induced by avian winged helix proteins.";
RL Cancer Res. 57:123-129(1997).
CC -! FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -! SUBCELLULAR LOCATION: Nuclear.
CC -! SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U37274; AAC60066.1; -
CC HSSP; Q63245; 2HFH.
CC TRANSFAC; T02495.
CC InterPro; IPR001766; Fork_head.
CC Pfam; PF00250; Fork_head; 1.
CC PRINTS; PR00053; FORKHEAD.
CC SMART; SM00339; FH; 1.
CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC PROSITE; PS00039; FORK_HEAD_3; 1.
CC DNA-binding; Nuclear protein; Transcription regulation.
CC FT DOMAIN 67 70 POLY-ALA.
CC FT DOMAIN 80 91 POLY-GLY.
CC FT DOMAIN 100 106 POLY-ALA.
CC FT DNA_BIND 117 211 FORK-HEAD.
CC SQ SEQUENCE 394 AA; 40995 MW; 3244436B9E31899 CRC64;
Query Match 38.1%; Score 68.5; DB 1; Length 394;
Best Local Similarity 46.3%; Pred. No. 2.3;
Matches 19; Conservative 2; Mismatches 9; Indels 11; Gaps 1;

QY 7 FOALAA-----RAGGGGGGIEGPTLRQALAAARA 36

ID	LACL_NEUCR	STANDARD;	PRT;	619 AA.
AC	P06811;			
AD	01-JAN-1988 (Rel. 06, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)			
DE	(Urushiol oxidase) (Laccase allele OR).			
GN	LACC.			
OS	Neurospora crassa			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OX	Sordariales; Sordariaceae; Neurospora.			
OC	NCBI-TaxID=5141;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RA	MEDLINE=88087214; PubMed=2961749;			
RT	Germann U.A., Mueller G., Hunziker P.E., Lerch K.;			
RT	"Characterization of two allelic forms of Neurospora crassa laccase:			
RT	Amino- and carboxyl-terminal processing of a precursor.";			
RL	J. Biol. Chem. 263:885-896(1988).			
RN	[2]			
RP	SEQUENCE OF 379-619 FROM N.A.			
RX	MEDLINE=87067412; PubMed=2947240;			
RA	Germann U.A., Lerch K.;			
RT	"Isolation and partial nucleotide sequence of the laccase gene from			
RT	Neurospora crassa: amino acid sequence homology of the protein to			
RT	human ceruloplasmin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).			
CC	-1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED			
CC	PRODUCTS (PROBABLE).			
CC	-1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzenesemiquinone + 2			
CC	H(2)O.			
CC	-1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU			
CC	CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE			
CC	3 OR COUPLED BINUCLEAR (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Secreted (Potential).			
CC	-1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.			
CC	-1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-			
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstat-			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in			
CC	modified and this statement is not removed. Usage by and for commer-			
CC	entities requires a license agreement (See http://www.isb-sib.ch/annou			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M14554; AAA33590.1; -			
DR	EMBL; M18333; AAA33591.1; -			
DR	PIR; A28523; KSNCLQ.			
DR	PIR; A29762; A29762.			
DR	InterPro: IPR001117; Cu-oxidase.			
DR	InterPro: IPR002355; MultiCu_oxidase2.			
DR	Pfam: PF00394; Cu-oxidase; 3.			
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.			
DR	PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.			
KW	Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;			
KW	Glycoprotein; Repeat.			
FT	SIGNAL 1			
FT	CHAIN 1 21			
FT	PROPEP 22 49			
FT	CHAIN 50 606			
FT	PROPEP 607 619			
FT	DOMAIN 84 207			
FT	DOMAIN 216 373			
FT	DOMAIN 431 566			
FT	DOMAIN 144 144			
FT	METAL 146 146			
FT	METAL 189 189			
FT	METAL 191 191			
FT	METAL 477 477			
FT	METAL 480 480			
FT	METAL 482 482			
FT	METAL 548 548			
FT	METAL 549 549			

RESULT 5
LAC1_NEUCR

RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
CC CHROMOSOME. IT IS ALSO INVOLVED IN DNA RECOMBINATION AND REPAIR
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SSB FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 137.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39923; AAB53120.1; ALT_FRAME.
CC EMBL; AL022118; CAAL7953.1; -
CC EMBL; AL583926; CAC32216.1; -
CC Leproma; ML2684; -
CC HSSP; P02339; LEVG; -
CC InterPro; IPR000424; SSB.
CC Pfam; PF00436; SSB; 1.
CC PROSITE; PS00735; SSB_1; FALSE_NEG.
CC PROSITE; PS00736; SSB_2; FALSE_NEG.
CC DNA-binding; DNA repair; DNA replication; Complete proteome.
CC DOMAIN 124 133
CC FT SEQUENCE 168 AA; 17700 MW; 077C62E430623658 CRC64;
CC SQ
CC
CC Query Match 34.7%; Score 62.5; DB 1; Length 168;
CC Best Local Similarity 56.0%; Pred. No. 4.3;
CC Matches 14; Conservative 3; Mismatches 5; Gaps 1;
CC
CC Qy 3 GPTLRQAL-----AARAGGGGGGG 22
CC ||||| . :|:||||||| |
CC Db 107 GPSLRYATKVNKASRSRGGGGGFG 131
CC
CC RESULT 15
CC ID TTF1_CAVPO STANDARD; PRT; 112 AA.
CC AC P97273;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE Thyroid transcription factor 1 (Thyroid nuclear factor 1) (TTF-1)
CC DE (Homeobox protein NKX-2.1) (Fragment).
CC GN TTF1 OR TTF1.
CC OS Cavia porcellus (Guinea pig).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
CC OX NCBI_TaxID=10141;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-DUNKIN-HARTLEY; TISSUE=Lung;
CC RA Yuan H.T., Bingle C.D.;
CC RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
CC THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
CC ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U82718; AAB40921.1; -
CC HSSP; P23441; 1FTT.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC SMART; SM00389; Hox; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS00071; HOMEBOX_2; 1.
CC KW Transcription regulation; Activator; Homeobox; DNA-binding;
CC Nuclear protein.
CC FT NON_TER 1 1
CC FT DNA_BIND 1 60 HOMEBOX.
CC FT NON_TER 112 112
CC SQ SEQUENCE 112 AA; 12723 MW; AEEAEDF06905F9DB CRC64;
CC
CC Query Match 34.4%; Score 62; DB 1; Length 112;
CC Best Local Similarity 52.0%; Pred. No. 3.3;
CC Matches 13; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
CC
CC Qy 7 RQALAAAGGGGGGGGIEGPTLRQA 31
CC :|:|:||||||| |
CC Db 66 QOOLQODSGGGGGGAGCPQQQA 90
CC
CC RESULT 16
CC ID TTF1_CANFA STANDARD; PRT; 371 AA.
CC AC P43698;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE Thyroid transcription factor 1 (Thyroid nuclear factor 1) (TTF-1)
CC DE (Homeobox protein NKX-2.1).
CC GN TTF1 OR TTF1.
CC OS Canis familiaris (Dog).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CC OX NCBI_TaxID=9615;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA MEDLINE=96034516; PubMed=7589789;
CC RA van Renterghem P.H.G., Drenier S., Vassar G., Christophe J.;
CC RT "Study of TTF-1 gene expression in dog thyrocytes in primary
CC culture";
CC RL Mol. Cell. Endocrinol. 112:83-93(1995).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
CC THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
CC ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: THYROID, LUNG AND CNS.
CC -!- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X77910; CAA54868.1; -
CC HSSP; P23441; 1FTT.
CC TRANSFAC; T02098; -
CC

us-09-422-838c-29.rsp

Wed Oct 9 10:30:02 2002

```

DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR Transcription regulation; Activator; Homeobox; DNA-binding;
KW Nuclear protein.
FT DNA_BIND 161 220 HOMEBOX.
FT DOMAIN 234 243 POLY-GLY.
FT DOMAIN 246 253 POLY-GLN.
FT DOMAIN 294 303 POLY-ALA.
FT SEQUENCE 371 AA; 38539 MW; 3F16CEB8E562604D7 CRC64;
SQ
Query Match 34.4%; Score 62; DB 1; Length 371;
Best Local Similarity 52.0%; Pred. No. 9.6;
Matches 13; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
QY 7 RQALARAGGGGGGGEGTTLRQA 31
Db 226 QOQLQDSDSGGGGGGAGCPQQQQA 250

RESULT 17
TTF1_HUMAN STANDARD; PRT; 371 AA.
ID P43699; O14955; O14954;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor 1 (thyroid nuclear factor 1) (TTF-1)
DE (Homeobox protein NKX-2.1).
DE TTF1 OR TTF1 OR NKX2A.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95226464; PubMed=7711080;
RA Salardi A., Tassi V., de Filippis V., Civitareale D.;
RT "Cloning and sequence analysis of human thyroid transcription factor
RT 1.";
RL Biochim. Biophys. Acta 1261:307-310(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95229626; PubMed=7713914;
RA Ikeda K., Clark J.C., Shaw-White J.R., Stahlman M.T., Boutell C.J.,
RA Whitsett J.A.;
RT "Gene structure and expression of human thyroid transcription
RT factor-1 in respiratory epithelial cells.";
RL J. Biol. Chem. 270:8108-8114(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95229626; PubMed=7713914;
RA Oguni H., Kimura S.;
RT "Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases."
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=95229626; PubMed=7713914;
RA Handan H., Liu H., Delemos R., Minoo P.;
RT "Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases."
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=95229626; PubMed=7713914;
RA Handan H., Liu H., Jones C., Delemos R., Minoo P.;
RT "Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases."
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=95229626; PubMed=7713914;
RA Endo T., Ohno M., Nakazato M.;
RT "Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases."
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

-!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
-!- SUBCELLULAR LOCATION: Nuclear.
-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
PRODUCED BY ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: THYROID AND LUNG.
-!- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
EMBL; X82850; CRA58053.1; -
EMBL; U19816; AAC50125.1; -
EMBL; U19756; AAA86099.1; -
EMBL; U43203; AAA89066.1; ALT_INIT.
EMBL; U33749; AAB52381.1; -
EMBL; D50740; BAA23529.1; -
EMBL; D50739; BAA23527.1; -
HSP; P23441; 1FTT
TRANSFAC; T00857; -
MIM; 600635; -
InterPro; IPR001356; Homeobox.
Pfam; PF00046; homeobox; 1.
PRINTS; PR00024; HOMEBOX.
SMART; SM00389; HOX; 1.
PROSITE; PS00027; HOMEBOX_1; 1.
PROSITE; PS00071; HOMEBOX_2; 1.
Transcription regulation; Activator; Homeobox; DNA-binding;
KW Nuclear protein; Alternative splicing.
FT DNA_BIND 161 220 HOMEBOX.
FT DOMAIN 234 243 POLY-GLY.
FT DOMAIN 246 253 POLY-GLN.
FT DOMAIN 294 303 POLY-ALA.
FT VARSPLIC 112 125 GWYGANPDRFPAL -> V (IN ISOFORM 2).
FT CONFLICT 49 61 P -> H (IN REF. 6).
FT CONFLICT 61 61 H -> P (IN REF. 6).
FT CONFLICT 158 158 S -> T (IN REF. 6).
FT CONFLICT 161 161 R -> G (IN REF. 6).
FT CONFLICT 226 227 QQ -> HE (IN REF. 5).
FT SEQUENCE 371 AA; 38596 MW; 5F1E3B40A1BBD862 CRC64;
SQ
Query Match 34.4%; Score 62; DB 1; Length 371;
Best Local Similarity 52.0%; Pred. No. 9.6;
Matches 13; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
QY 7 RQALARAGGGGGGGEGTTLRQA 31
Db 226 QOQLQDSDSGGGGGGAGCPQQQQA 250

RESULT 18
EVX2_HUMAN STANDARD; PRT; 476 AA.
ID EVX2_HUMAN
AC Q03828;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Homeobox even-skipped homolog protein 2 (EVX-2).
DE EVX2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=98066765; PubMed=9403061;
RA Ernstson S., Betz R., Lagercrantz S., Larsson C., Erickson S.,
RA Cederberg A., Carlsson P., Enerbaeck S.;
RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
RT expressed human forkhead gene that maps to chromosome 1p32-p34.";
RL Genomics 46:78-85(1997).
RN [2]
RP REVISIONS.
RA Enerbaeck S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF042832; AAC15421.1; -.
DR HSSP: Q63245; 2HFH.
DR TRANSFAC: T02485; -.
DR MIM: 602211; -.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
KW DOMAIN 90 94 POLY-ALA.
FT DOMAIN 101 104 POLY-ALA.
FT DOMAIN 126 217 FORK-HEAD.
FT DOMAIN 247 250 POLY-ALA.
FT DOMAIN 296 306 POLY-ALA.
FT DOMAIN 398 409 POLY-GLY.
FT DOMAIN 421 426 POLY-GLY.
FT DOMAIN 442 445 POLY-ALA.
SQ SEQUENCE 497 AA; 49007 MW; EAAF498D216BE019 CRC64;
Query Match 34.4%; Score 62; DB 1; Length 497;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
QY 4 PT--LRQALAAAGGGGGGGG 22
Db ||||| ||||| |||||
385 PTALLRQGLKTAGGGAGGGG 405

RESULT 20
HB9_HUMAN STANDARD; PRT; 401 AA.
AC P50219;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Homeobox protein HB9.
LN HLXB9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RL MEDLINE=94327547; PubMed=7914194;
RA Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;

```


Wed Oct 9 10:30:02 2002

```

DE      Hypothetical protein TP0967.
GN      TP0967.
OS      Treponema pallidum.
OC      Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX      NCBI_TaxID=160;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NICHOLS;
RX      MEDLINE=9832770; PubMed=965876;
RA      Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA      Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA      Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA      Khakhal H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA      McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA      Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA      Venter J.C.;
RT      "Complete genome sequence of Treponema pallidum, the syphilis
RT      spirochete."
RL      Science 281:375-388(1998).
CC      -|- SIMILARITY: BELONGS TO THE TP096X FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AE001264; AAC65925.1; -
CC      TIGR: TP0967; -
CC      Hypothetical protein; Complete proteome.
KW      Hypothetical 152 161 POLY-GLY.
FT      DOMAIN 152 161
SQ      SEQUENCE 517 AA; 56597 MW; E224976333989DF6 CRC64;

Query Match      33.9%; Score 61; DB 1; Length 517;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY      3 CPTLRQALARAGGGGGG 22
DB      141 GMTVTQPNAGAGGGGGG 160
      | | | | |
      | | | | |

RESULT 25
ZIN_HUMAN
ID      ZIN_HUMAN          STANDARD;          PRT;          753 AA.
AC      Q9NRL3;
DT      01-MAR-2002 (Rel. 41, Created)
DT      01-MAR-2002 (Rel. 41, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Zinedin.
GN      ZIN.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20347911; PubMed=10748158;
RA      Castets F., Rakitina T., Gaillard S., Moqrigh A., Mattei M.-G.,
RA      Monneron A.;
RT      "Zinedin, SG2NA, and striatin are calmodulin-binding, WD repeat
RT      proteins principally expressed in the brain."
RL      J. Biol. Chem. 275:19970-19977(2000).
RN      [2]
RP      SEQUENCE OF 402-753 FROM N.A.
RC      TISSUE=Muscle;
RA      Strausberg R.;
RL      Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases. MAY
CC      -|- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER.
CC      -|- FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.
CC      -|- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND (BY

```

CC SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.

CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC -!- CAUTION: The name "zinedin" probably originates from the name of

CC the famous soccer player from Marseille (Zinedine Zidane)!

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF212940; AAF29527.1; -

DR EMBL; BC004910; AAH04910.1; -

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 7.

DR PRINTS; PR00320; GPROTEINBRPT.

DR SMART; SM00320; WD40; 6.

DR PROSITE; PS00678; WD_REPEATS.1; 1.

DR PROSITE; PS50082; WD_REPEATS.2; 4.

DR PROSITE; PS50294; WD_REPEATS_REGION; 1.

DR Calmodulin-binding; Repeat; WD repeat; Coiled coil.

DR DOMAIN 69 136

FT CALMODULIN-BINDING (POTENTIAL).

FT DOMAIN 165 182

FT REPEAT 436 475

FT REPEAT 489 528

FT REPEAT 542 581

FT REPEAT 587 628

FT REPEAT 635 674

FT REPEAT 677 716

FT REPEAT 723 752

FT SITE 71 79

FT DOMAIN 402 404

FT CONFLICT 402 404

FT POLY-ALA.

FT LAD -> GTR (IN REF. 2).

FT SEQUENCE 753 AA; 80581 MW; 4DA016A8FF7EDB5E CRC64;

Query Match 33.9%; Score 61; DB 1; Length 753;

Best Local Similarity 48.4%; Pred. No. 22;

Matches 15; Conservative 2; Mismatches 8; Indels 6; Gaps 1;

QY 3 GPTLRQALAA-----RAGGGGGGGIEGPT 27

DB 27 GPTGAAPVSAPAGPGAGPGAGGGGGGSGPT 57

|||||::| | | | | | | |

DB 27 GPTGAAPVSAPAGPGAGPGAGGGGGGSGPT 57

RESULT 26

SXL_CERCA

ID SXL_CERCA STANDARD; PRT; 348 AA.

AC 061374;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sex-lethal protein homolog (CCSLX).

GN SXL.

OS Ceratitis capitata (Mediterranean fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta.

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Tephritidae; Tephritidae; Ceratitis.

OX NCBI_TaxID=7213;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN-BENAKIO;

RA MEDLINE=98171464; PubMed=9502730;

RA Saccone G., Peluso I., Artiano D., Giordano E., Bopp D., Polito L.C.;

RT "The Ceratitis capitata homologue of the Drosophila sex-determining

RT gene Sex-lethal is structurally conserved, but not sex-specifically

RL regulated".

RL Development 125:1495-1500(1998).

CC -!- FUNCTION: UNKNOWN; APPARENTLY NOT INVOLVED IN SOMATIC SEX

CC DETERMINATION.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS: ADULT-SPECIFIC ISOFORMS

CC A1, A2, A3, A4, AND EMBRYO-SPECIFIC ISOFORMS E1, E2 AND E3 (SHOWN

CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS OF BOTH SEXES. ALSO

CC EXPRESSED IN THE PROGENITOR CELLS OF THE GERMLINE.

CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF026145; AAC38968.1; -

DR HSP; P19339; 1SXL.

DR InterPro; IPR000504; RRM.

DR Pfam; PF00076; rrm; 2.

DR PRINTS; PR00961; HUDSXLRNA.

DR SMART; SM00360; RRM; 2.

DR PROSITE; PS50102; RRM; 2.

DR PROSITE; PS00030; RRM_RNP.1; 1.

DR RNA-binding; Repeat; Nuclear protein; Alternative splicing.

DR DOMAIN 1 27

FT GLY/ASN-RICH DOMAIN.

FT DOMAIN 110 188

FT RNA-BINDING (RRM) 1.

FT DOMAIN 196 276

FT RNA-BINDING (RRM) 2.

FT DOMAIN 68 75

FT POLY-GLY.

FT DOMAIN 95 99

FT POLY-GLY.

FT DOMAIN 293 311

FT POLY-GLY.

FT DOMAIN 312 316

FT POLY-PRO.

FT VARSPLIC 37 44

FT MISSING (IN ISOFORM A1).

FT SEQUENCE 348 AA; 37188 MW; CABA3DA5C2C8874A CRC64;

Query Match 33.3%; Score 60; DB 1; Length 348;

Best Local Similarity 83.3%; Pred. No. 14;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGP 26

DB 301 GGGGGGGGMP 312

|||||::| | | | | | | |

DB 301 GGGGGGGGMP 312

RESULT 27

BR3A_MOUSE

ID BR3A_MOUSE STANDARD; PRT; 421 AA.

AC P17208;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Brain-specific homeobox/POU domain protein 3A (BRN-3A) (BRN-3.0).

GN POU4F1 OR BRN3A OR BRN3 OR BRN-3.

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

[1]

RP SEQUENCE FROM N.A.

RC MEDLINE=94215319; PubMed=8162704;

RA Theil T., Zechner U., Klett C., Adolph S., Moeroy T.;

RT "Chromosomal localization and sequences of the murine Brn-3 family of

RT developmental control genes".

RL Cytogenet. Cell Genet. 66:267-271(1994).

[2]

RP SEQUENCE OF 286-401 FROM N.A.

RC STRAIN-T6/TW1; Tissue-Testis;

RX MEDLINE=90221898; PubMed=1970171;

RA Goldborough A., Ashworth A., Willison K.;

RT "Cloning and sequencing of POU-boxes expressed in mouse testis".

RL Nucleic Acids Res. 18:1634-1634(1990).

CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR WHICH MAY PLAY A ROLE IN

CC THE REGULATION OF SPECIFIC GENE EXPRESSION WITHIN A SUBSET OF

CC NEURONAL LINEAGES. MAY PLAY A ROLE IN DETERMINING OR MAINTAINING

THE IDENTITIES OF A SMALL SUBSET OF VISUAL SYSTEM NEURONS.

-1- SUBCELLULAR LOCATION: Nuclear.

-1- TISSUE SPECIFICITY: BRAIN, PERIPHERAL SENSORY NERVOUS SYSTEM AND RETINA. IN THE ADULT NERVOUS SYSTEM BRN-3.0 PREDOMINATES IN THE MEDIAL HABENULA, SUPERFICIAL GRAY OF THE SUPERIOR COLLICULUS, RED NUCLEUS, MESENCEPHALIC NUCLEUS OF THE TRIGEMINAL GANGLION, NUCLEUS AMBIGUUS, INFERIOR OLIVARY NUCLEUS, AND PERIPHERAL SENSORY GANGLIA.

-1- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS TO CLASS-4 POU.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: S69350; AAB30577.2; -
 EMBL: X51959; CAA36218.1; -
 PIR: S09237; S09237.
 HSP: P14859; ILOC.
 MGD: MGI:102525; Pou4f1.
 InterPro: IPR001356; Homeobox.
 PIR: PF00046; homeobox; 1.
 PIR: PF00157; pou; 1.
 PRODOM: PD000583; pou; 1.
 SMART: SM00389; HOX; 1.
 SMART: SM00352; POU; 1.
 PROSITE: PS00027; HOMEBOX_1; 1.
 PROSITE: PS00035; POU_1; 1.
 PROSITE: PS00465; POU_2; 1.
 PROSITE: PS50071; HOMEBOX_2; 1.
 Transcription regulation; Nuclear protein; DNA-binding; Homeobox; Developmental protein.

DOMAIN 57 66 POU-IV BOX.
 FT DOMAIN 100 108 POLY-HIS.
 FT DOMAIN 130 187 GLY-RICH.
 FT DOMAIN 205 258 ALA-RICH.
 FT DOMAIN 265 339 POU.
 FT DOMAIN 357 416 HOMEBOX.
 FT SEQUENCE 421 AA; 42781 MW; 34EC99D789EBE939 CRC64;

Query Match 33.3%; Score 60; DB 1; Length 421;
 Best Local Similarity 53.3%; Pred. No. 17;
 Matches 16; Conservative 2; Mismatches 6; Indels 6; Gaps 2;

QY 1 TEGPTLRQALAAAGG---GGGGGGIEGP 26
 | 1: ||||| ||||| :||
 Db 120 ISSPSL--ALMAGAGAGAGAGGGGAGHDGP 147

RESULT 28
 KICJ_MOUSE STANDARD; PRT; 569 AA.
 AC P02535; P08731;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (56 kDa cytokeratin)
 DE (Keratin, type I cytoskeletal 59 kDa).
 GN KRT10 OR KRT1-10.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85207552; PubMed=2581944;
 RA Krieg T.M., Schafer M.P., Cheng C.K., Filipula D., Flaherty P.,

Steinert P.M., Roop D.R.;
 "Organization of a type I keratin gene. Evidence for evolution of intermediate filaments from a common ancestral gene.";
 J. Biol. Chem. 260:5867-5870(1985).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE=83192464; PubMed=6189855;
 Steinert P.M., Rice R.H., Roop D.R., Trus B.L., Steven A.C.;
 "Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments.";
 Nature 302:794-800(1983).
 -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS. KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
 -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC) (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
 -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L00193; AAA39391.1; -
 EMBL: M10081; AAA39391.1; JOINED.
 EMBL: V00830; CAA24214.1; -
 PIR: A02940; KRMSE1.
 PIR: S07330; S07330.
 HSP: P10968; LWGC.
 MGD: MGI:96685; Krt1-10.
 InterPro: IPR001664; IF.
 InterPro: IPR002957; Keratin_I.
 PIR: PF00038; filament; 1.
 PRINTS: PR01248; TYPE1KERATIN.
 PROSITE: PS00226; IF; 1.
 Intermediate filament; Coiled coil; Keratin.

INIT_MET 0 HEAD.
 FT DOMAIN 1 142 ROD.
 FT DOMAIN 143 453 TAIL.
 FT DOMAIN 454 569 COIL 1A.
 FT DOMAIN 143 178 LINKER 1.
 FT DOMAIN 179 199 COIL 1B.
 FT DOMAIN 200 291 LINKER 12.
 FT DOMAIN 292 314 COIL 2.
 FT DOMAIN 315 453 STUTTER.
 FT SITE 395 395 GLY/SER-RICH.
 FT DOMAIN 452 564 S -> C (IN REF. 2).
 FT CONFLICT 5 5 S -> F (IN REF. 2).
 FT CONFLICT 24 24 S -> F (IN REF. 2).
 FT CONFLICT 28 28 Y -> L (IN REF. 2).
 FT CONFLICT 38 38 E -> G (IN REF. 2).
 FT CONFLICT 41 41 AG -> GS (IN REF. 2).
 FT CONFLICT 104 105 MISSING (IN REF. 2).
 FT CONFLICT 110 110 SY -> GC (IN REF. 2).
 FT CONFLICT 121 122 Q -> R (IN REF. 2).
 FT CONFLICT 137 137 WYKHGNSQ -> VVREARQLKP (IN REF. 2).
 FT CONFLICT 148 148 KSDLEL -> QSVLEL (IN REF. 2).
 FT CONFLICT 178 187 H -> L (IN REF. 2).
 FT CONFLICT 263 268 E -> A (IN REF. 2).
 FT CONFLICT 284 284 EGRYCV -> VESILR (IN REF. 2).
 FT CONFLICT 353 353 GGSHGGS -> CGRGGG (IN REF. 2).
 FT CONFLICT 394 399 S -> G (IN REF. 2).
 FT CONFLICT 508 514 H -> R (IN REF. 2).
 FT CONFLICT 523 523 S -> G (IN REF. 2).
 FT CONFLICT 531 531 S -> G (IN REF. 2).
 FT CONFLICT 534 543 S -> G (IN REF. 2).
 FT CONFLICT 543 548 GQ -> RR (IN REF. 2).
 FT CONFLICT 547 556 KS -> SGT (IN REF. 2).
 FT CONFLICT 555 569 AA; 57711 MW; EEC59D4D8FF6484D CRC64;
 SEQUENCE

Query Match 33.3%; Score 60; DB 1; Length 796;

1155 GGGGGGGVLPNLSQ II/4

1133

1155 GGGGGGGVLFNLSQ 11/4

11/4

us-09-422-838c-29.rsp

Wed Oct 9 10:30:02 2002

Search completed: October 9, 2002, 09:00:16
Job time : 5.3831 secs

.....

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-29
Perfect score: 180
Sequence: 1 IEPTLRQALAAAGGGGGGIEPTLRQALAA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mbc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	42.2	360	10 Q9LGC9	Q9LGC9 oryza sativ
2	73	40.6	253	10 Q943K0	Q943K0 oryza sativ
3	72	40.0	496	2 Q9AD76	Q9AD76 streptomyce
4	70	38.9	66	12 Q9LBC5	Q9LBC5 spodoptera
5	70	38.9	137	10 Q9M6A1	Q9M6A1 catharanthu
6	70	38.9	160	10 Q9M699	Q9M699 catharanthu
7	70	38.9	612	4 Q9E270	Q9E270 homo sapien
8	69.5	38.6	165	2 Q9AF15	Q9AF15 mycobacteri
9	69	38.3	381	10 Q9LD54	Q9LD54 oryza sativ
10	68.5	38.1	339	10 Q9S270	Q9S270 arabidopsis
11	68.5	38.1	796	5 Q27258	Q27258 drosophila
12	68.5	38.1	797	5 Q9V7U9	Q9V7U9 drosophila
13	68.5	38.1	806	5 Q9E828	Q9E828 drosophila
14	68.5	38.1	867	10 Q42696	Q42696 chlamydomon
15	68	37.8	474	4 Q96SQ2	Q96SQ2 homo sapien
16	68	37.8	529	10 Q9ASE5	Q9ASE5 oryza sativ

17	68	37.8	688	4 Q9BYD8	Q9BYD8 homo sapien
18	68	37.8	689	4 Q96JG7	Q96JG7 homo sapien
19	68	37.8	752	4 Q96L34	Q96L34 homo sapien
20	67	37.2	113	10 Q942U6	Q942U6 oryza sativ
21	67	37.2	1186	5 Q6I080	Q6I080 acanthamoeb
22	66	36.7	416	5 Q9W254	Q9W254 drosophila
23	66	36.7	532	10 Q19476	Q19476 caenorhabdi
24	66	36.7	532	10 Q9FWD7	Q9FWD7 oryza sativ
25	65	36.1	155	5 Q9GND8	Q9GND8 drosophila
26	65	36.1	155	5 Q9GND8	Q9GND8 drosophila
27	65	36.1	156	5 Q9GND8	Q9GND8 drosophila
28	65	36.1	156	5 Q9GND8	Q9GND8 drosophila
29	65	36.1	157	5 Q9GND8	Q9GND8 drosophila
30	65	36.1	158	5 Q9GND8	Q9GND8 drosophila
31	65	36.1	159	5 Q9GND8	Q9GND8 drosophila
32	65	36.1	159	5 Q9GND8	Q9GND8 drosophila
33	65	36.1	161	5 Q9GND8	Q9GND8 drosophila
34	65	36.1	161	5 Q9GND8	Q9GND8 drosophila
35	65	36.1	162	5 Q9GND8	Q9GND8 drosophila
36	65	36.1	163	5 Q9GND8	Q9GND8 drosophila
37	65	36.1	163	5 Q9GND8	Q9GND8 drosophila
38	65	36.1	163	5 Q9GND8	Q9GND8 drosophila
39	65	36.1	163	5 Q9GND8	Q9GND8 drosophila
40	65	36.1	164	5 Q9GND8	Q9GND8 drosophila
41	65	36.1	164	5 Q9GND8	Q9GND8 drosophila
42	65	36.1	165	5 Q9GND8	Q9GND8 drosophila
43	65	36.1	165	5 Q9GND8	Q9GND8 drosophila
44	65	36.1	165	5 Q9GND8	Q9GND8 drosophila
45	65	36.1	165	5 Q9GND8	Q9GND8 drosophila

ALIGNMENTS

RESULT 1

Q9LGC9
ID Q9LGC9 PRELIMINARY; PRT; 360 AA.
AC Q9LGC9
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-OCT-2001 (TREMREL. 18, Last annotation update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0462H08.19";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002525; BAB07996.1;
DR InterPro; IPR000571; zf-CCCH;
DR Pfam; PF00642; zf-CCCH; 4.
DR SMART; SM00356; Znf_C3H1; 4.
SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 42.2%; Score 76; DB 10; Length 360;
Best Local Similarity 56.0%; Pred No. 0.66;
Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 IEPTLRQALAAAGGGGGGIEPTLRQALAA 25

Db 26 LEGPMWRMGLGGGGGGGGGGG 50

RESULT 2

Q943K0 PRELIMINARY; PRT; 253 AA.
ID Q943K0

Q943KO; 01-DEC-2001 (TrEMBLrel. 19, Created)
01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
P0039A07.6 PROTEIN.
P0039A07.6
Oryza sativa (Rice).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoidae; Oryzeae; Oryza.
NCBI_TaxID=4530;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0039A07," to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003235; BAB64100.1; A963166CE5F97B2B CRC64;
SQ SEQUENCE 253 AA; 25568 MW; A963166CE5F97B2B CRC64;

Query Match 40.6%; Score 73; DB 10; Length 253;
Best Local Similarity 55.6%; Pred. No. 1;
Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 3 GPTLRQAALARAGGGGGGIEGPTLR 29
||| : ||| |||||| | :
DB 80 GPTVGVRVAYRAGAGGGGGPRGFALK 106

RESULT 3
Q9AD76 PRELIMINARY; PRT: 496 AA.

ID Q9AD76
AC Q9AD76; 2001 (TrEMBLrel. 17, Created)
DT DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN SKI3.27.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacterlia; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomycetes.
OX NCBI_Taxid=1902;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb streptomycetes coelicolor A3(2) chromosome";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL512667; CAC21636.2; -
DR InterPro; IPR003838; DUF214.
DR Pfam; PF02687; DUF214; 1.
SQ SEQUENCE 496 AA; 49548 MW; 54E110C4F86231A4 CRC64;

Query Match 40.0%; Score 72; DB 2; Length 496;
Best Local Similarity 48.7%; Pred. No. 2.5;
Matches 19; Conservative 2; Mismatches 8; Indels 10; Gaps 2;

OY 4 PTLRQAALARAGGGGGGGGIEGPTLRQAALA 33
||| : ||| ||| ||| ||| |||

Wed Oct 9 10:30:03 2002

us-09-422-838c-29.rspt

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
 CC EMBL: AP001383; BAA92520.1; -.
 CC EMBL: AP001080; BAA90348.1; -.
 DR InterPro: IPR001993; Mitoch_carrier.
 DR InterPro: IPR002067; Mitoch_carrier.
 DR Pfam: PF00153; mito_carr; 3.
 DR PRINTS: PR00926; MITOCARRIER.
 DR PROSITE: PS00215; MITOCH_CARRIER; 2.
 DR Inner membrane; Mitochondrion; Transmembrane; Transport.
 KW SEQUENCE 381 AA; 40761 MW; F3A0E3CEB950778 CRC64;
 SQ SEQUENCE 381 AA; 40761 MW; F3A0E3CEB950778 CRC64;
 Query Match 38.3%; Score 69; DB 10; Length 381;
 Best Local Similarity 51.9%; Pred. No. 4.1;
 Matches 14; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
 QY 4 PTLRQALAAAGGGGGGIEGPTLRQ 30
 DB 23 PHARRALALRVGGGGGPAFASLTVRE 49
 I:||||| I:||||| I:|
 RESULT 10
 Q9SZ70 PRELIMINARY; PRT; 339 AA.
 AC Q9SZ70;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PUTATIVE DNA-BINDING PROTEIN.
 GN F16J13.120 OR AT4G12050.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;
 RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL049638; CAB40946.1; -.
 DR EMBL: AL161533; CAB78248.1; -.
 KW DNA-binding.
 SQ SEQUENCE 339 AA; 35785 MW; A7E3A9C73B6BBE21 CRC64;
 Query Match 38.1%; Score 68.5; DB 10; Length 339;
 Best Local Similarity 38.5%; Pred. No. 4.1;
 Matches 15; Conservative 8; Mismatches 13; Indels 3; Gaps 1;
 QY 1 IEPTLRQALAAAGGGGGGIEGPTLRQALAA 36
 DB 268 LEEDQMTPVQGGGGGGGGGPPPMQQAAMA 306
 I:||||| I:||||| I:|
 RESULT 11
 Q27258 PRELIMINARY; PRT; 796 AA.
 ID Q27258
 AC Q27258;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PSI PROTEIN.
 GN PSI OR EG:EG0003.2 OR CG8912.
 CC Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95172378; PubMed=7867926;
 RA Siebel C.W., Admon A., Rio D.C.;
 RA "Soma-specific expression and cloning of PSI, a negative regulator of
 RT P element pre-mRNA splicing.";
 RL Genes Dev. 9:269-283(1995).
 DR EMBL: S75665; AAB50636.1; -.
 DR FlyBase: FBgn0014870; PSI.
 DR InterPro: IPR004087; KH.
 DR InterPro: IPR004088; KH_TYPE_1.
 DR InterPro: IPR002173; PfKB.
 DR Pfam: PF00013; KH-domain; 4.
 DR SMART: SM00322; KH; 4.
 DR PROSITE: PS00884; KH_TYPE_1; 4.
 DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
 SQ SEQUENCE 796 AA; 81588 MW; 4B9423E033677EF5 CRC64;
 Query Match 38.1%; Score 68.5; DB 5; Length 796;
 Best Local Similarity 34.0%; Pred. No. 9.4;
 Matches 16; Conservative 7; Mismatches 11; Indels 13; Gaps 1;
 QY 3 GPTLRQAL-----AARAGGGGGGIEGPTLRQALAA 36
 DB 62 GPESKRSLSPEYANNSSNMSSGGGGGGGGGGGSGITQAIQAA 108
 I:||||| I:||||| I:|
 RESULT 12
 Q9V7U9 PRELIMINARY; PRT; 797 AA.
 ID Q9V7U9
 AC Q9V7U9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PSI PROTEIN.
 GN PSI OR EG:EG0003.2 OR CG8912.
 CC Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostek A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

RESULT 13
096828
ID 096828 PRELIMINARY; PRT; 806 AA.
AC AC 096828;
DT DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE EG:EG0003.2 PROTEIN.
GN PSI OR EG:EG0003.2 OR CG8912.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL "Sequencing the distal x chromosome of Drosophila melanogaster.";
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031863; CAA21318.1; -.
DR FlyBase; FBgn0014870; Psi.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH_TYPE_1.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00013; KH-domain; 4.
DR SMART; SM00322; KH; 4.
DR PROSITE; PS00084; KH_TYPE_1; 4.
DR PROSITE; PS00583; PFKB_KINASE_1; UNKNOWN_1.
SQ SEQUENCE 806 AA; 84180 MW; CC9AG367110CA8FE CRC64;

```

```

RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027619; BAB55238.1; -.
SQ SEQUENCE 474 AA; 51313 MW; 0BCA301518F20DED CRC64;

Query Match 37.8%; Score 68; DB 4; Length 474;
Best Local Similarity 45.0%; Pred. No. 6.4;
Matches 18; Conservative 1; Mismatches 15; Indels 6; Gaps 1;

QY 3 GPTLR-----QALAAAGGGGGGEGPTLRQALAA 36
   |||:| | | | | | | | | | | | | | | | | |
Db 284 GSTIRTFHGGVDRRAGGGGGVQNGPPASPTLAHEA 323

RESULT 16
Q9ASE5 PRELIMINARY; PRT; 529 AA.
AC Q9ASE5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE P0456F08.14 PROTEIN.
GN P0456F08.14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 1, PAC
   clone:P0456F08."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002901; BAB39414.1; -.
DR InterPro; IPR002937; Amino_oxidase.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF01593; Amino_oxidase; 1.
SQ SEQUENCE 529 AA; 55981 MW; 0A5DA55CD076D24 CRC64;

Query Match 37.8%; Score 68; DB 10; Length 529;
Best Local Similarity 55.2%; Pred. No. 7.2;
Matches 16; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 6 LRQALAAAGGGGGGEGPTLRQALAA 34
   |||:| | | | | | | | | | | | | | | | | |
Db 151 LRAYQAARSAGGGGGGKLEEVDEALLA 179

RESULT 17
Q9BYD8 PRELIMINARY; PRT; 688 AA.
AC Q9BYD8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAP/MICROTUBULE AFFINITY-REGULATING KINASE LIKE 1.
GN MARKL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX Kato T., Satoh S., Okabe H., Kichara O., Ono K., Kihara C.,
   Tanaka T., Tsunoda T., Yamaoka Y., Nakamura Y.;
   "Isolation of a novel human gene, MARKL1, homologous to MARK3 and its
   involvement in hepatocellular carcinogenesis."
RL Neoplasia 3:4-9(2001).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

```

```

DR EMBL; AB049127; BAB39380.1; -.
DR HSPF; Q63450; IA06.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR SMART; SM00165; UBA; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 688 AA; 75261 MW; A03B5A7943ACD086 CRC64;

Query Match 37.8%; Score 68; DB 4; Length 688;
Best Local Similarity 45.0%; Pred. No. 9.2;
Matches 18; Conservative 1; Mismatches 15; Indels 6; Gaps 1;

QY 3 GPTLR-----QALAAAGGGGGGEGPTLRQALAA 36
   |||:| | | | | | | | | | | | | | | | | |
Db 562 GSTIRTFHGGVDRRAGGGGGVQNGPPASPTLAHEA 601

RESULT 18
Q96JG7 PRELIMINARY; PRT; 689 AA.
AC Q96JG7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1860 PROTEIN (FRAGMENT).
GN KIAA1860.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
   "Prediction of the coding sequences of unidentified human genes. XX.
   The complete sequences of 100 new cDNA clones from brain which code
   for large proteins in vitro."
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058763; BAB47489.1; -.
DR NON_TER;
FT NON_TER;
SQ SEQUENCE 689 AA; 75449 MW; 439B11FD33D78B34 CRC64;

Query Match 37.8%; Score 68; DB 4; Length 689;
Best Local Similarity 45.0%; Pred. No. 9.2;
Matches 18; Conservative 1; Mismatches 15; Indels 6; Gaps 1;

QY 3 GPTLR-----QALAAAGGGGGGEGPTLRQALAA 36
   |||:| | | | | | | | | | | | | | | | | |
Db 563 GSTIRTFHGGVDRRAGGGGGVQNGPPASPTLAHEA 602

RESULT 19
Q96L34 PRELIMINARY; PRT; 752 AA.
AC Q96L34;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MARK4 SERINE/THREONINE PROTEIN KINASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```



```

RN RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=97262070; PubMed=9108484;
RX Drewes G., Ebner A., Preuss U., Mandelkow E.M., Mandelkow E.;
RT "MARK, a novel family of protein kinases that phosphorylate
FT microtubule-associated proteins and trigger microtubule disruption.";
RL Cell 89:297-308(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Drewes G., Mandelkow E.M.;
RT "MARK4, homologue of MARK1, MARK2 and MARK3.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057448; AAL23683.1;
SQ KINASE: Serine/threonine-protein kinase.
KW SEQUENCE 752 AA; 82519 MW; 4B430FFD2B150E7A CRC64;

Query Match 37.8%; Score 68; DB 4; Length 752;
Best Local Similarity 45.0%; Pred. No. 10;
Matches 18; Conservative 1; Mismatches 15; Indels 6; Gaps 1;

QY 3 GPTLR-----QALARAGGGGGGIEGPTLRQALAA 36
| | | | | | | | | | | | | | | | | | | |
Db 562 GSTIRTFHGQVRRDRAGGGGGVQNGPPASPTLAHEA 601

RESULT 20
Q94206 PRELIMINARY; PRT; 113 AA.
AC Q94206;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DE P0506E04.26 PROTEIN.
GN P0506E04.26
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone:P0506E04.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003272; BAB67948.1;
SQ SEQUENCE 113 AA; 11708 MW; 26D9B2C86935BC0B CRC64;

Query Match 37.2%; Score 67; DB 10; Length 113;
Best Local Similarity 59.1%; Pred. No. 2.1;
Matches 13; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 15 GGGGGGGGGTGGTPTLRQALAA 36
| | | | | | | | | | | | | |
Db 85 GGGGGGGGGTGGTPTLRQALAA 106

RESULT 21
O61080 PRELIMINARY; PRT; 1186 AA.
AC O61080;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
GN MYOSIN IC HEAVY CHAIN.
OS MICHC.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed=3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
and non-myosin-like sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99079990;
RA Wang Z.Y., Wang F., Sellers J.R., Korn E.D., Hammer J.A. III;
RT "Analysis of the regulatory phosphorylation site in Acanthamoeba
myosin IC by using site-directed mutagenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:15200-15205(1998).
DR EMBL; AF051353; AAC98089.1;
DR HSSP; P08799; 1MND.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
SQ SEQUENCE 1186 AA; 129459 MW; E37AD44A685803A6 CRC64;

Query Match 37.2%; Score 67; DB 5; Length 1186;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 14; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 8 QALAAAGGGGGGGGIEGPTLRQALAA 35
| | | | | | | | | | | | | |
Db 938 QILGAKGGGGGGGGRGGPSPGAVSPR 965

RESULT 22
Q9W254 PRELIMINARY; PRT; 416 AA.
AC Q9W254;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE QKRS8E-2 PROTEIN.
GN QKRS8E-2 OR QKRS8E-2 OR CG5821.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer G., Helt G., Chao M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

Wed Oct 9 10:30:03 2002

```

Ra Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Ra Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Ra Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Ra Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
Ra Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
Ra Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Ra Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Ra Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Ra Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Ra Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
Ra Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Ra Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Ra Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Ra Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ra Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Ra Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
Ra Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
Ra "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL EMBL; AE003457; AAF46845.1; -.
DR FlyBase; FBgn0022985; qkr38E-2.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH_TYPE_1.
DR Pfam; PF00013; KH-domain; 1.
DR SMART; SM00322; KH; 1.
DR PROSITE; PS50084; KH_TYPE_1; 1.
SQ SEQUENCE 416 AA; 44144 MW; 095FF2F5D87D2A5 CRC64;

Query Match 36.7%; Score 66; DB 5; Length 416;
Best Local Similarity 59.3%; Pred. No. 9.4;
Matches 16; Conservative 1; Mismatches 4; Indels 6; Gaps 1;

QY 5 TLRLQ-----ALAAAGGGGGGGGGG 25
DB 253 TLRLQSQSNAASGAGGGGGGGG 279

RESULT 23
Q19476 PRELIMINARY; PRT; 500 AA.
AC Q19476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F15B9.5 PROTEIN.
GN Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z78013; CAB01420.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 500 AA; 53946 MW; 1416327086FE7CF6 CRC64;

Query Match 36.7%; Score 66; DB 5; Length 500;
Best Local Similarity 56.5%; Pred. No. 11;
Matches 13; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 GPTLRQALARPAGGGGGGGG 25

```

```

Db 429 GSMGLRFLNRGGGGGGGGG 451

RESULT 24
Q9FWD7 PRELIMINARY; PRT; 532 AA.
AC Q9FWD7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEICAL 57.3 KDA PROTEIN.
GN OSJNB0018B10.3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNB0018B10 genomic sequence."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC051634; AAG13445.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003595; AAA_subfam.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW ATP-binding; Hypothetical protein.
SQ SEQUENCE 532 AA; 57290 MW; 7EA6CC4F4C6734C5 CRC64;

Query Match 36.7%; Score 66; DB 10; Length 532;
Best Local Similarity 51.9%; Pred. No. 12;
Matches 14; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGGGGGPTLRQALARA 36
DB 1 MTPREGGGGGGGVGLVAYALAVVA 27

RESULT 25
Q9GP74 PRELIMINARY; PRT; 155 AA.
AC Q9GP74;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NONA PROTEIN (FRAGMENT).
GN NONA.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ALAI;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
Drosophila littoralis and D. virilis."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304359; CAC20082.1; -.
DR FlyBase; FBgn0043410; Diit\NonA.
DR NON_TER; 1
FT NON_TER 155
SQ SEQUENCE 155 AA; 15362 MW; 5DCE33593769FC57 CRC64;

Query Match 36.1%; Score 65; DB 5; Length 155;
Best Local Similarity 80.0%; Pred. No. 4.6;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 8 QALAAAGAGGGGGG 22
Db 63 QAFRAGGGGGGGG 77

RESULT 26
Q9GND8 PRELIMINARY; PRT; 155 AA.
AC Q9GND8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NONA PROTEIN (NO ON OR OFF TRANSIENT A) (FRAGMENT).
GN NONA.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RU1;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L11;
RA Huttunen S., Vieira J., Hoikkala A.;
RT "Levels and patterns of nucleotide variability and homopolymer length
RT variation at the nonA gene in Drosophila virilis group species.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304365; CAC20038.1; -
DR EMBL; AJ304365; CAC20038.1; -
DR EMBL; AJ304305; CAC20038.1; -
DR EMBL; AJ304308; CAC20031.1; -
DR EMBL; AJ304311; CAC20034.1; -
DR EMBL; AJ304317; CAC20040.1; -
DR EMBL; AJ304320; CAC20043.1; -
DR EMBL; AJ304323; CAC20046.1; -
DR EMBL; AJ304326; CAC20049.1; -
DR EMBL; AJ304329; CAC20052.1; -
DR EMBL; AJ304332; CAC20055.1; -
DR EMBL; AJ304335; CAC20058.1; -
DR EMBL; AJ304338; CAC20061.1; -
DR EMBL; AJ304341; CAC20064.1; -
DR EMBL; AJ304344; CAC20067.1; -
DR EMBL; AJ304353; CAC20076.1; -
DR EMBL; AJ304356; CAC20079.1; -
DR EMBL; AY012599; AG48869.1; -
DR FlyBase; FBgn0043410; Dlit\nonA.
FT NON_TER 1
FT NON_TER 155
SQ SEQUENCE 155 AA; 15336 MW; 5DCE33592CC84657 CRC64;

Query Match 36.1%; Score 65; DB 5; Length 155;
Best Local Similarity 80.0%; Pred. No. 4.6;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 QALAAAGAGGGGGG 22
Db 63 QAFRAGGGGGGGG 77

RESULT 27
Q9GP73 PRELIMINARY; PRT; 156 AA.
AC Q9GP73;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NONA PROTEIN (FRAGMENT).
GN NONA.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RU3, OU4, AND SA3;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304368; CAC20091.1; -
DR EMBL; AJ304314; CAC20037.1; -
DR EMBL; AJ304347; CAC20070.1; -
DR FlyBase; FBgn0043410; Dlit\nonA.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 15393 MW; 7B6202DB1A7DCD51 CRC64;

Query Match 36.1%; Score 65; DB 5; Length 156;
Best Local Similarity 80.0%; Pred. No. 4.7;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 QALAAAGAGGGGGG 22
Db 63 QAFRAGGGGGGGG 77

RESULT 28
Q9GNB7 PRELIMINARY; PRT; 156 AA.
AC Q9GNB7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NONA PROTEIN (FRAGMENT).
GN NONA.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RU3, OU4, AND SA3;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304368; CAC20091.1; -
DR EMBL; AJ304314; CAC20037.1; -
DR EMBL; AJ304347; CAC20070.1; -
DR FlyBase; FBgn0043410; Dlit\nonA.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 15393 MW; 7B6202DB1A7DCD51 CRC64;

Query Match 36.1%; Score 65; DB 5; Length 156;
Best Local Similarity 80.0%; Pred. No. 4.7;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 QALAAAGAGGGGGG 22
Db 63 QAFRAGGGGGGGG 77

RESULT 29
Q9GP77 PRELIMINARY; PRT; 157 AA.
AC Q9GP77;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NONA PROTEIN (FRAGMENT).
GN NONA.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RU1;
RA Huttunen S., Campesan S., Hoikkala A.;
RT "Intra- and interspecific nucleotide variation at the nonA gene in
RT Drosophila littoralis and D. virilis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304362; CAC20085.1; -
DR EMBL; AJ304362; CAC20085.1; -
DR FlyBase; FBgn0043410; Dlit\nonA.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 15428 MW; 808202C5D5413BF0 CRC64;

Query Match 36.1%; Score 65; DB 5; Length 156;
Best Local Similarity 80.0%; Pred. No. 4.7;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 QALAAAGAGGGGGG 22
Db 63 QAFRAGGGGGGGG 77

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 16.1874 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-30

Perfect score: 193

Sequence: 1 IEQPTLQWLARAGGKGGGIEGPTLQWLARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	36	21	AA17301
2	193	100.0	36	21	AA17301
3	186	96.4	36	21	AA16963
4	186	96.4	36	21	AA17293
5	186	96.4	36	21	AA16963
6	186	96.4	41	21	AA169528
7	186	96.4	42	21	AA17281
8	186	96.4	42	21	AA17282
9	186	96.4	42	21	AA17308
10	186	96.4	42	21	AA169530
11	186	96.4	60	21	AA17311

12	186	96.4	269	21	AA16960	TMP-TMP-Fc protein
13	186	96.4	269	21	AA16963	Human IgG1 Fc TWP
14	185	95.9	36	21	AA17303	TPO-mimetic peptid
15	185	95.9	36	21	AA17307	TPO-mimetic peptid
16	185	95.9	36	21	AA169524	Thrombopoietin mim
17	182	94.3	36	21	AA17306	Thrombopoietin mim
18	182	94.3	36	21	AA169526	TPO-mimetic peptid
19	182	94.3	268	21	AA169529	Thrombopoietin mim
20	181.5	94.0	39	21	AA16959	FC-TMP-TMP protein
21	181	93.8	40	21	AA17304	TPO-mimetic peptid
22	177.5	92.0	35	21	AA17302	TPO-mimetic peptid
23	175.5	90.9	37	21	AA17292	TPO-mimetic peptid
24	175	90.7	38	21	AA17294	TPO-mimetic peptid
25	174.5	90.4	39	21	AA17295	TPO-mimetic peptid
26	173	89.6	39	21	AA17305	TPO-mimetic peptid
27	171	88.6	42	21	AA17296	TPO-mimetic peptid
28	164.5	85.2	33	21	AA17291	TPO-mimetic peptid
29	160	82.9	36	21	AA17290	TPO-mimetic peptid
30	160	82.9	36	21	AA17298	TPO-mimetic peptid
31	160	82.9	36	21	AA17299	TPO-mimetic peptid
32	158	81.9	32	21	AA169521	Cyclic or linear t
33	158	81.9	36	21	AA17289	TPO-mimetic peptid
34	158	81.9	36	21	AA17300	TPO-mimetic peptid
35	151.5	78.5	36	21	AA169522	Linear thrombopoie
36	145	75.1	31	21	AA17288	TPO-mimetic peptid
37	144	74.6	30	21	AA17287	TPO-mimetic peptid
38	144	74.6	32	21	AA17297	TPO-mimetic peptid
39	144	74.6	32	21	AA169520	Thrombopoietin mim
40	138.5	71.8	34	21	AA169527	Thrombopoietin mim
41	132	68.4	29	21	AA17286	TPO-mimetic peptid
42	131.5	68.1	28	21	AA17285	TPO-mimetic peptid
43	129.5	67.1	29	21	AA16970	TPO-mimetic peptid
44	129.5	67.1	31	21	AA16973	TPO-mimetic peptid
45	125.5	65.0	31	21	AA16974	TPO-mimetic peptid
			29	21	AA16971	TPO-mimetic peptid

ALIGNMENTS

RESULT 1
ID AA17301 standard; Peptide; 36 AA.
XX
AC AA17301;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:357.
XX

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antitumor; thrombolytic; VEGF;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0103371.
XX
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
XX WPI; 2000-350702/30.

Wed Oct 9 10:30:04 2002

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases

XX Example 1; Page 321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, (L4)f-P4
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;
 Query Match 100.0%; Score 193; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 9.1e-17;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAARA 36
 |||||||||||||||||||||||||||||||||||||
 Db 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAARA 36

RESULT 2
 AAB169523
 ID AAB169523 standard; peptide; 36 AA.

XX AAB169523;
 XX 04-SEP-2000 (first entry)

Thrombopoietin mimetic peptide compound 4.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14

FT Peptide 15..22 /label= TMP_1

FT Modified-site 18 /label= linker

FT Peptide 23..36 /note= "optionally modified by bromoacetyl or PEG"

FT Peptide /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX

(AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-nTMP-2],
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A, F, M, or K; X-7 = R or Q; X-8 = O, N,
 CC or E; X-9 = W, Y or F; X-10 = L, I, V, L, F, G, S, or Q; X-11 = A, I, V,
 CC L, F, S, T, K, H, or E; X-12 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 100.0%; Score 193; DB 21; Length 36;

Best Local Similarity 100.0%; Pred. No. 9.1e-17;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAARA 36
 |||||||||||||||||||||||||||||||||||||

Db 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAARA 36

RESULT 3

AAB16963

ID AAB16963 standard; Protein; 36 AA.

XX AAB16963;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide TMP-TMP SEQ ID NO:14.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases -

XX

XX Disclosure; Page 190; 608pp; English.

XX

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX where P1, P2, P3, and P4 = are each independently sequences of

XX pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX independently linkers; and a, b, c, d, e, and f = are each

XX 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX activities. DNAs, vectors and host cells from the present invention can

XX be used for producing pharmaceutical compositions. The compositions are

XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

XX The use of an Fc domain (rather than a Fab domain) can provide a longer

XX half-life or incorporate functions such as Fc receptor binding, protein

XX A binding, complement fixation, and possibly placental transfer. AAA69443

XX to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

XX sequences used in the exemplification of the present invention.

XX

XX Sequence 36 AA;

XX

Query Match 96.4%; Score 186; DB 21; Length 36;

Best Local Similarity 97.2%; Pred. No. 6.4e-16;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36

|||||

Db 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36

|||||

RESULT 4

AAB17293

ID AAB17293 standard; Peptide; 36 AA.

XX

XX AAB17293;

XX

XX 31-OCT-2000 (first entry)

XX

XX TPO-mimetic peptide sequence SEQ ID NO:349.

XX

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CPTA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX

XX Synthetic.

XX

XX WO200024782-A2.

XX

XX 04-MAY-2000.

XX

XX 25-OCT-1999; 99WO-US25044.

XX

XX 23-OCT-1998; 98US-0105371.

XX

XX 22-OCT-1999; 99US-0428082.

XX

XX (AMGE-) AMGEN INC.

XX

XX Feige U, Liu C, Cheetham J, Boone TC;

XX

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases -

XX

XX Example 1; Page 318; 608pp; English.

XX

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX where P1, P2, P3, and P4 = are each independently sequences of

XX pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX independently linkers; and a, b, c, d, e, and f = are each

XX 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX activities. DNAs, vectors and host cells from the present invention can

XX be used for producing pharmaceutical compositions. The compositions are

XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

XX The use of an Fc domain (rather than a Fab domain) can provide a longer

XX half-life or incorporate functions such as Fc receptor binding, protein

XX A binding, complement fixation, and possibly placental transfer. AAA69443

XX to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

XX sequences used in the exemplification of the present invention.

XX

XX Sequence 36 AA;

XX

Query Match 96.4%; Score 186; DB 21; Length 36;

Best Local Similarity 97.2%; Pred. No. 6.4e-16;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36

|||||

Db 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36

|||||

RESULT 5

AAV96525

ID AAV96525 standard; Peptide; 36 AA.

XX

XX AAV96525;

XX

XX 04-SEP-2000 (first entry)

XX

XX Thrombopoietin mimetic peptide compound 6.

XX

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;

XX anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;

XX immunosuppressive; anti-inflammatory; linker.

XX

XX Synthetic.

XX

XX Key Location/Qualifiers

XX Modified-site 1 /note= "optionally linked to an Fc molecule"

XX Peptide 1..14 /label= TMP_1

XX Peptide 15..18 /label= linker

XX Peptide 19..32 /label= TMP_2

XX Modified-site 32 /note= "optionally linked to an Fc molecule"

XX

XX WO200024770-A2.

XX

XX 04-MAY-2000.

XX

XX 22-OCT-1999; 99WO-US24834.

XX

XX 23-OCT-1998; 98US-0105348.

XX

(ANGE-) AMGEN INC.

Liu C, Feige U, Cheetham J;

WPI; 2000-365108/31.

Thrombopoietic peptides which activate mpl receptors and increase the production of platelets or platelet precursors, useful for treatment of diseases which involve thrombocytopenia

Claim 16; Page 62; 91pp; English.

A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1).nTMP_2], is to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2, X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and X_1-X_1_4, X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A; X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N, or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V, T, V, N, Q or G; X_1_2 = A, I, V, L, F, T, R, E, or G; L_1 = linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

Sequence 36 AA;

Query Match 96.4%; Score 186; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 6.4e-16;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 IEGPTLQWLARAGCGGGGIEGPTLRQWLAAARA 36
 ||||||||||||||| |||||||||||||||
 Db 1 IEGPTLRQWLARAGCGGGGIEGPTLRQWLAAARA 36

RESULT 6

AA196528

ID AA196528 standard; peptide; 41 AA.

XX AC AA196528;

XX DT 04-SEP-2000 (first entry)

XX DE Thrombopoietin mimetic peptide compound 9.

XX KW Thrombopoietin; mimetic; TMP; TPQ; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX OS Synthetic.

Key	Location/Qualifiers
FT Modified-site	1 /note= "optionally linked to an Fc molecule"
FT Peptide	6..19 /label= TMP_1
FT Peptide	20..27 /label= linker
FT Peptide	28..41 /label= TMP_2

XX WO200024770-A2.

XX PD 04-MAY-2000.

XX PD 22-OCT-1999; 99WO-US24834.

XX

[illegible][illegible]

Wed Oct 9 10:30:04 2002

DR WPI: 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX

XX Example 2; Page 327; 608pp; English.

PS The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ

Query Match 96.4%; Score 186; DB 21; Length 42;

Best Local Similarity 97.2%; Pred. No. 7.5e-16;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARAGCGGGGIEGPTLRQWLARA 36

DB 7 IEPTLRQWLARAGCGGGGIEGPTLRQWLARA 42

RESULT 10

AA196530

ID AA196530 standard; Protein: 42 AA.

AC AA196530;

XX

XX 04-SEP-2000 (first entry)

XX

DE Thrombopoietin mimetic peptide.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;

KW megakaryocyte, production; anti-human immunodeficiency virus; anti-HIV;

KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX

OS Synthetic.

XX

XX WO200024770-A2.

PN

XX

XX 04-MAY-2000.

PD

XX

XX 22-OCT-1999; 99WO-US24834.

PF

XX

XX 23-OCT-1998; 98US-0105348.

PR

XX

XX (AMGE-) AMGEN INC.

PA

XX

XX Liu C, Feige U, Cheetham J;

PI

XX

XX WPI: 2000-365108/31.

DR

XX

XX N-PSDB; AAA29225.

DR

XX

XX Thrombopoietic peptides which activate mpl receptors and increase the

PT production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

XX

Example 2A; Page 48; 91pp; English.

PS Overlapping oligonucleotides were used to construct a synthetic

XX gene encoding a thrombopoietin mimetic peptide (TMP), which

CC was then fused in-frame to the Fc region of the human IgG1 chain (see

CC AAY96529). A compound which binds to an mpl receptor comprising a TMP

CC dimer joined by a linker [TMP₁-(L₁)-TMP₂], is new. TMP₁ and TMP₂

CC are amino acid sequences varying from at least 10 to 14 residues in

CC length comprising X₂-X₁-L₁, X₂-X₁-L₂, X₂-X₁-L₃, X₂-X₁-L₄, and

CC X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and X₁-X₁-L₄. X₁ = L, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S;

CC X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N, or E; X₉ = W, Y or F;

CC X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, L, V, L, F, S, T, K, H, or E;

CC X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K, T, V, N, Q or G; X₁₄ =

CC A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino

CC acids; and n = 0 or 1. The compounds bind to and activate the c-mpl

CC receptor which mediates the activity of endogenous thrombopoietin. The

CC TMPs are useful for increasing the production of platelets or platelet

CC precursors (e.g. megakaryocytes) in a mammal, which is useful for

CC treatment of diseases which involve thrombocytopenia, e.g. aplastic

CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus

CC associated ITP, and systemic lupus erythematosus.

XX

SQ

Sequence 42 AA;

Query Match 96.4%; Score 186; DB 21; Length 42;

Best Local Similarity 97.2%; Pred. No. 7.5e-16;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARAGCGGGGIEGPTLRQWLARA 36

DB 7 IEPTLRQWLARAGCGGGGIEGPTLRQWLARA 42

RESULT 11

AAB17311

ID AAB17311 standard; Peptide: 60 AA.

XX AAB17311;

XX

XX 31-OCT-2000 (first entry)

XX

DE Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.

XX

XX Modified peptide; therapeutic agent; fusion: Fc domain; cancer;

KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

OS Homo sapiens.

OS Synthetic.

XX WO200024782-A2.

PN

XX

XX 04-MAY-2000.

PD

XX

XX 25-OCT-1999; 99WO-US25044.

PF

XX

XX 23-OCT-1998; 98US-0105371.

PR

XX

XX 22-OCT-1999; 99US-0428082.

PR

XX

XX (AMGE-) AMGEN INC.

PA

XX

XX Feige U, Liu C, Cheetham J, Boone TC;

PI

XX

XX WPI: 2000-350702/30.

DR

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

PT

XX PS Example 2; Page 331; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

CC FC domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions can

CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. Protein

CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 60 AA;

Query Match 96.4%; Score 186; DB 21; Length 60;

Best Local Similarity 97.2%; Pred. No. 1.1e-15;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAARA 36

|||||

Db 2 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAARA 37

RESULT 12

AAB16960

ID AAB16960 standard; Protein; 269 AA.

XX AC AAB16960;

XX DT 31-OCT-2000 (first entry)

XX DE Tmp-TMP-Fc protein sequence SEQ ID NO:10.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; cancer;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Homo sapiens.

XX QS Synthetic.

XX PN WO200024782-A2.

XX XX 04-MAY-2000.

XX PD 25-OCT-1999; 99WO-US25044.

XX PF 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX XX (AMGE-) AMGEN INC.

XX PA Feige U, Liu C, Cheetham J, Boone TC;

XX PI WPI; 2000-350702/30.

XX DR N-PSDB; AAB69446.

XX XX Novel composition of matter comprising an Fc domain and

XX PT pharmacologically active peptides, useful for treating cancer and

XX PT autoimmune diseases.

XX PS Example 2; Page 185-186; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

CC FC domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions can

CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. Protein

CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 269 AA;

Query Match 96.4%; Score 186; DB 21; Length 269;

Best Local Similarity 97.2%; Pred. No. 5.2e-15;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAARA 36

|||||

Db 2 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAARA 37

RESULT 13

AAY96531

ID AAY96531 standard; Protein; 269 AA.

XX AC AAY96531;

XX DT 04-SEP-2000 (first entry)

XX DE Human IgG1 Fc TMP fusion protein.

XX KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; platelet;

XX KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;

XX KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX OS Homo sapiens.

XX PN WO200024770-A2.

XX PD 04-MAY-2000.

XX XX 22-OCT-1999; 99WO-US24834.

XX PF 23-OCT-1998; 98US-0105348.

XX PR (AMGE-) AMGEN INC.

XX PA Liu C, Feige U, Cheetham J;

XX PI WPI; 2000-365108/31.

XX DR N-PSDB; AAA29229.

XX XX Thrombopoietic peptides which activate mpl receptors and increase the

XX PT production of platelets or platelet precursors, useful for treatment of

XX PT diseases which involve thrombocytopenia

XX PS Example 2A; Page 49-50; 91pp; English.

XX CC A compound which binds to an mpl receptor comprising a thrombopoietin

XX CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],

XX CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 $(X1)a-F1-(x2)b$, where: F1 = an Fc domain; x1 and x2 = are each independently selected from -L(L)-C-P1-(L2)-d-P2,
-L(L)-C-P1-(L2)d-P2-(L3)e-p3, Or -L(L)c-P1-(L2)-d-(L3)e-p3-(L4)f-p4
where p1, p2, p3, and p4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently

1

10

XX	Key	Location/Qualifiers
FT	Modified-site	1
FT	Peptide	/note= "optionally linked to an Fc molecule"
FT	Peptide	1..14
FT	Peptide	/label= TMP_1
FT	Peptide	15..18
FT	Peptide	/label= linker
FT	Peptide	19..32
XX		/label= TMP_2
XX	WO200024770-A2.	
XX	04-MAY-2000.	
XX	22-OCT-1999;	99WO-US24834.
XX	23-OCT-1998;	98US-0105348.
XX	(AMGE-) AMGEN INC.	
XX	Liu C, Feige U, Cheetham J;	
XX	WPT; 2000-365108/31.	
XX	Thrombopoietic peptides which activate mpl receptors and increase the	
PT	production of platelets or platelet precursors, useful for treatment of	
PT	diseases which involve thrombocytopenia	
XX	Claim 16; Page 62; 91pp; English.	
XX	A compound which binds to an mpl receptor comprising a thrombopoietin	
CC	mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-nTMP_2],	
CC	is new. TMP_1 and TMP_2 are amino acid sequences varying from at least	

RESULT 19	
AAB16959	AAB16959 standard; Protein; 268 AA.
ID	AA16959
XX	AA16959;
XX	31-OCT-2000 (first entry)
XX	FC-TMP-TMP protein sequence SEQ ID NO:8.
XX	Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW	autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW	immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW	MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW	vascular endothelial growth factor; matrix metalloproteinase;
KW	asthma; thrombosis; pharmaceutical.
XX	Homo sapiens.
OS	Synthetic.
XX	WO200024782-A2.
PN	04-MAY-2000.
XX	25-OCT-1999; 99WO-US25044.
XX	23-OCT-1998; 98US-0105371.
PR	23-OCT-1999; 99US-0428082.
XX	(AMGE-) AMGEN INC.
PA	Feige U, Liu C, Cheetham J, Boone TC;
XX	WPI; 2000-350702/30.
DR	N-PSDB; AAA69445.
XX	Novel composition of matter comprising an Fc domain and
PT	pharmacologically active peptides, useful for treating cancer and
PT	autoimmune diseases -
XX	Example 2; Page 182-183; 608pp; English.
PS	The present invention describes composition of matter (1) comprising an
XX	Fc domain, pharmacologically active peptides, and linkers. Where (1) is
CC	(X1)a-Fl-(X2)b, where: Fl = an Fc domain; X1 and X2 = are each
CC	independently selected from -(L1)c-Pl, -(L1)c-Pl-(L2)d-P2,

CC (L1)c-P1-(L2)d-P2-(L3)e-P*3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX Sequence 268 AA;
 SQ

Query Match 94.3%; Score 182; DB 21; Length 268;
 Best Local Similarity 97.1%; Pred. No. 1.6e-14;
 Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 IEPTLQWLAAARAGGKGEGPTLQWLAAAR 35
 DB 234 IEPTLQWLAAARAGGKGEGPTLQWLAAAR 268

RESULT 20
 AAB17304
 ID AAB17304 standard; Peptide; 39 AA.
 AC AAB17304;
 XX

31-OCT-2000 (first entry)
 DE TPO-mimetic peptide sequence SEQ ID NO:360.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

Synthetic.

WO2000024782-A2.

04-MAY-2000.

25-OCT-1999; 99WO-US25044.

23-OCT-1998; 98US-0105371.

22-OCT-1999; 99US-0428082.

(AMGE-) AMGEN INC.

Feige U, Liu C, Cheetham J, Boone TC;
 WPI: 2000-350702/30.

Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX

Example 1; Page 323; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P*3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX Sequence 39 AA;
 SQ

Query Match 94.0%; Score 181.5; DB 21; Length 39;
 Best Local Similarity 92.3%; Pred. No. 2.4e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 IEPTLQWLAAARAGGKGEGPTLQWLAAAR 36
 DB 1 IEPTLQWLAAARAGGKGEGPTLQWLAAAR 39

RESULT 21
 AAB17302
 ID AAB17302 standard; Peptide; 40 AA.
 AC AAB17302;
 XX

31-OCT-2000 (first entry)
 DE TPO-mimetic peptide sequence SEQ ID NO:358.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

Synthetic.

WO2000024782-A2.

04-MAY-2000.

25-OCT-1999; 99WO-US25044.

23-OCT-1998; 98US-0105371.

22-OCT-1999; 99US-0428082.

(AMGE-) AMGEN INC.

Feige U, Liu C, Cheetham J, Boone TC;
 WPI: 2000-350702/30.

Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX

Example 1; Page 322; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P*3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

Wed Oct 9 10:30:04 2002

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 40 AA;
 SQ Query Match 93.8%; Score 181; DB 21; Length 40;
 Best Local Similarity 90.0%; Pred. No. 2.9e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 TEGPTLROWLAARAGGK---GGGGEGPTLROWLAARA 36
 |||||
 DB 1 TEGPTLROWLAARAGGKBRACGGGGEGPTLROWLAARA 40
 |||||

RESULT 22
 AAB17292
 ID AAB17292 standard; Peptide: 35 AA.
 XX
 AC AAB17292;
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:348.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 XX WO200024782-A2.
 XX
 XX 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -

XX Example 1; Page 317-318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are

CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 35 AA;
 SQ Query Match 92.0%; Score 177.5; DB 21; Length 35;
 Best Local Similarity 97.2%; Pred. No. 6.7e-15;
 Matches 35; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 TEGPTLROWLAARAGGGGGEGPTLROWLAARA 36
 |||||
 DB 1 TEGPTLROWLAARAGG-GGGGEGPTLROWLAARA 35
 |||||

RESULT 23
 AAB17294
 ID AAB17294 standard; Peptide: 37 AA.
 XX
 AC AAB17294;
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:350.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 XX WO200024782-A2.
 XX
 XX 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -

XX Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX
 SQ Sequence 37 AA;

Query Match 90.9%; Score 175.5; DB 21; Length 37;
 Best Local Similarity 94.6%; Pred. No. 1.2e-14;
 Matches 35; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1 IEGPTLRQWLAAARA--GGGKGGGIEGPTLRQWLAAARA 36
 Db 1 IEGPTLRQWLAAARAAGGGGGGGGIEGPTLRQWLAAARA 37

RESULT 24
 AAB17295
 ID AAB17295 standard; Peptide; 38 AA.
 XX
 AC AAB17295;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:351.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO2000024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;
 PI WPI: 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX
 SQ Sequence 38 AA;

Query Match 90.7%; Score 175; DB 21; Length 38;
 Best Local Similarity 92.1%; Pred. No. 1.5e-14;
 Matches 35; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

OY 1 IEGPTLRQWLAAARA--GGGKGGGIEGPTLRQWLAAARA 36
 Db 1 IEGPTLRQWLAAARAAGGGGGGGGIEGPTLRQWLAAARA 38

RESULT 25
 AAB17305
 ID AAB17305 standard; Peptide; 39 AA.
 XX
 AC AAB17305;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:361.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO2000024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;
 PI WPI: 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -

XX Example 1; Page 323; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443

sequence 34 AA;

XX 05

Query Match 88.6%; Score 171; DB 21; Length 34;
 Best Local Similarity 94.4%; Pred. No. 4e-14; Indels 0; Mismatches 0; Gaps 1;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 1;

OY 1 IEGPTLROWLAARAGGKGEGPTLROWLAARA 36
 |||||
 DB 1 IEGPTLROWLAARAGGKGEGPTLROWLAARA 34

RESULT 28
 AAB17290
 ID AAB17290 standard; Peptide; 33 AA.
 XX
 AC AAB17290;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:346.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX

XX WO200024782-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 25-OCT-1999; 99WO-US25044.
 XX
 XX PR 23-OCT-1998; 98US-0105371.
 XX PR 22-OCT-1999; 99US-0428082.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 XX DR WPI; 2000-350702/30.
 XX

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX PS Example 1; Page 317; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX SQ Sequence 33 AA;

Query Match 85.2%; Score 164.5; DB 21; Length 33;
 Best Local Similarity 91.7%; Pred. No. 9.1e-13;
 Matches 33; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Best Local Similarity 91.7%; Pred. No. 2.4e-13;
 Matches 33; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 1 IEGPTLROWLAARAGGKGEGPTLROWLAARA 36
 |||||
 DB 1 IEGPTLROWLAARAGGKGEGPTLROWLAARA 33

RESULT 29
 AAB17298
 ID AAB17298 standard; Peptide; 36 AA.
 XX
 AC AAB17298;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:354.
 XX

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX

XX WO200024782-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 25-OCT-1999; 99WO-US25044.
 XX
 XX PR 23-OCT-1998; 98US-0105371.
 XX PR 22-OCT-1999; 99US-0428082.
 XX
 XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 XX DR WPI; 2000-350702/30.
 XX

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX PS Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX SQ Sequence 36 AA;

Query Match 82.9%; Score 160; DB 21; Length 36;
 Best Local Similarity 91.7%; Pred. No. 9.1e-13;
 Matches 33; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 IEGPTLRWLAAARAGGGKGGGIEGPTLRQWLAARA 36

```
RESULT 30  
AAB17299  
ID AAB17299 standard; Peptide; 36 AA.  
XX  
XX AAB17299;  
XX AC  
XX DT 31-OCT-2000 (first entry)  
XX
```

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antitachytic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; cTfLA4; mimetic; IL-1; INF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asbma; thrombosis; pharmaceutical.

XX Synthetic.

XX PN WO200024782-A2.

XX
PD
04-MAY-2000.XX
PF
25-OCT-1999;

XX 23-OCT-1998; 98US-0105371.
PR

PR 22-OCT-1955, 5502
YX

PA (AMGE-) AMGEN INC.
YY

pi Feige U, Liu C, Guo X
 yy

DR 350702750.
XX
WPL; 2000

pharmacologically active peptides

PT autoimmune diseases

Example 1; Page 320 321; 322

CC The present invention describes
CC the domain, pharmacologically
CC

independently selected from

$$-(L1)c-p1-(L2)d-p2-(L3)e-p3$$

pharmacologically active peptides.

0 or 1, provided that at least one of the following conditions is satisfied:

CC activities. DNAs, vectors and

CC-90002 is useful for treating cancer,

CC half-life or incorporate in

CC to AAA69526 and AAB16953 to

XXXX

6030

Best Local Similarity 91.7%

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 : Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838C-30
Perfect score: 193
Sequence: 1 IEPTURQWLAAAGGGGGGIEGFTLROWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	76.5	39.6	25	2	US-08-764-640-231
2	76.5	39.6	25	3	US-09-244-298A-231
3	76.5	39.6	25	4	US-09-516-704-231
4	73	37.8	14	2	US-08-764-640-13
5	73	37.8	14	2	US-08-764-640-13
6	73	37.8	14	3	US-08-764-640-193
7	73	37.8	14	3	US-08-973-225-13
8	73	37.8	14	3	US-08-973-225-193
9	73	37.8	14	3	US-09-244-298A-13
10	73	37.8	14	4	US-09-516-704-13
11	73	37.8	14	4	US-09-516-704-193
12	73	37.8	15	2	US-08-764-640-17
13	73	37.8	15	2	US-08-764-640-185
14	73	37.8	15	3	US-08-973-225-17
15	73	37.8	15	3	US-08-973-225-185
16	73	37.8	15	3	US-09-244-298A-17
17	73	37.8	15	3	US-09-244-298A-185
18	73	37.8	15	4	US-09-516-704-17
19	73	37.8	15	4	US-09-516-704-185
20	73	37.8	16	2	US-08-764-640-18
21	73	37.8	16	2	US-08-764-640-194
22	73	37.8	16	3	US-08-764-640-232
23	73	37.8	16	3	US-08-973-225-18
24	73	37.8	16	3	US-08-973-225-194
25	73	37.8	16	3	US-08-973-225-220
26	73	37.8	16	3	US-09-244-298A-18
27	73	37.8	16	3	US-09-244-298A-194

28 73 37.8 16 3 US-09-244-298A-232 Sequence 232, App
29 73 37.8 16 4 US-09-516-704-18 Sequence 18, Appl
30 73 37.8 16 4 US-09-516-704-194 Sequence 194, App
31 73 37.8 16 4 US-09-516-704-232 Sequence 232, App
32 69 35.8 14 2 US-08-764-640-195 Sequence 195, App
33 69 35.8 14 2 US-08-764-640-195 Sequence 195, App
34 69 35.8 14 3 US-08-973-225-195 Sequence 195, App
35 69 35.8 14 3 US-08-973-225-195 Sequence 195, App
36 69 35.8 14 3 US-09-244-298A-195 Sequence 195, App
37 69 35.8 14 3 US-09-244-298A-195 Sequence 195, App
38 69 35.8 14 4 US-09-516-704-195 Sequence 195, App
39 69 35.8 14 4 US-09-516-704-195 Sequence 195, App
40 69 35.8 15 2 US-08-764-640-196 Sequence 196, App
41 69 35.8 15 2 US-08-764-640-200 Sequence 200, App
42 69 35.8 15 2 US-08-764-640-209 Sequence 209, App
43 69 35.8 15 2 US-08-764-640-215 Sequence 215, App
44 69 35.8 15 3 US-08-973-225-196 Sequence 196, App
45 69 35.8 15 3 US-08-973-225-200 Sequence 200, App

ALIGNMENTS

RESULT 1
US-08-764-640-231
: Sequence 231, Application US/08764640
: Patent No. 5869451
: Patent No. 5869451 5837683
: GENERAL INFORMATION:
: APPLICANT: Dower, William J.
: APPLICANT: Barrett, Ronald W.
: APPLICANT: Cwirla, Steven E.
: APPLICANT: Gates, Christian
: APPLICANT: Schatz, Peter J.
: APPLICANT: Balasubramanian, Palaniappan
: APPLICANT: Wagstrom, Christopher R.
: APPLICANT: Hendren, Richard W.
: APPLICANT: Deprince, Randolph B.
: APPLICANT: Podduturi, Surekha
: APPLICANT: Yin, Qun
: TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
: NUMBER OF SEQUENCES: 244
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Glaxo Wellcome
: STREET: Five Moore Drive, P.O. Box 13398
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,640
: FILING DATE: 11-DEC-1996
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Hrubiec, Robert T.
: REGISTRATION NUMBER: 36,392
: REFERENCE/DOCKET NUMBER: PK3281
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-248-1000
: INFORMATION FOR SEQ ID NO: 231:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 25 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FEATURE:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 37.8%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLROWLAARA 14

RESULT 5
US-08-764-640-13
Sequence 193, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193

Query Match 37.8%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLROWLAARA 14

RESULT 6
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherrill S.
APPLICANT: Mattheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A

Wed Oct 9 10:30:04 2002

us-09-422-838c-30.ra1

US-08-973-225-193

FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 37.8%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 7
US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 37.8%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-Dec-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 37.8%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 9
US-09-244-298A-193
; Sequence 193, Application US/09244298A

Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: linear
MOLECULE TYPE: peptide
US-09-244-298A-193
Query Match 37.8%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEQPTLRQWLAARA 14
Db 1 IEQPTLRQWLAARA 14
RESULT 10
US-09-516-704-13
Sequence 13, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13
Query Match 37.8%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEQPTLRQWLAARA 14
Db 1 IEQPTLRQWLAARA 14
RESULT 11
US-09-516-704-193
Sequence 193, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

us-09-422-838c-30.rai

Wed Oct 9 10:30:04 2002

```

; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hrubiec, Robert T.
;   REGISTRATION NUMBER: 36,392
;   REFERENCE/DOCKET NUMBER: PK3281
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 14 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: <Unknown>
;     MOLECULE TYPE: linear
;     TOPOLOGY: linear
;     SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193
;
; Query Match          37.8%; Score 73; DB 4; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.0053;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLROWLAARA 14
;   |||||
; DB 1 IEGPTLROWLAARA 14
;
; RESULT 12
; US-08-764-640-17
;   Sequence 17, Application US/08764640
;   Patent No. 5869451
;   Patent No. 5869451 5837683
;   GENERAL INFORMATION:
;     APPLICANT: Dower, William J.
;     APPLICANT: Barrett, Ronald W.
;     APPLICANT: Cwirla, Steven E.
;     APPLICANT: Gates, Christian
;     APPLICANT: Schatz, Peter J.
;     APPLICANT: Balasubramanian, Palaniappan
;     APPLICANT: Wagstrom, Christopher R.
;     APPLICANT: Hendren, Richard W.
;     APPLICANT: Deprince, Randolph B.
;     APPLICANT: Podduturi, Surekha
;     APPLICANT: Yin, Qun
;   TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;   TITLE OF INVENTION: RECEPTOR
;   NUMBER OF SEQUENCES: 244
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Glaxo Wellcome
;     STREET: Five Moore Drive, P.O. Box 13398
;     CITY: Research Triangle Park
;     STATE: NC
;     COUNTRY: USA
;     ZIP: 27709
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/764,640
;       FILING DATE: 11-DEC-1996
;       CLASSIFICATION: 514
;       ADDRESSEE: Glaxo Wellcome
;       STREET: Five Moore Drive, P.O. Box 13398
;       CITY: Research Triangle Park
;       STATE: NC
;       COUNTRY: USA
;       ZIP: 27709
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/764,640
;       FILING DATE: 11-DEC-1996
;       CLASSIFICATION: 514
;       ADDRESSEE: Glaxo Wellcome
;       STREET: Five Moore Drive, P.O. Box 13398
;       CITY: Research Triangle Park
;       STATE: NC
;       COUNTRY: USA
;       ZIP: 27709
;   INFORMATION FOR SEQ ID NO: 17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 15 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: linear
;       TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;     US-08-764-640-185
;
; Query Match          37.8%; Score 73; DB 2; Length 15;
; Best Local Similarity 100.0%; Pred. No. 0.0057;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLROWLAARA 14
;   |||||
; DB 2 IEGPTLROWLAARA 15
;
; RESULT 13
; US-08-764-640-185
;   Sequence 185, Application US/08764640
;   Patent No. 5869451
;   Patent No. 5869451 5837683
;   GENERAL INFORMATION:
;     APPLICANT: Dower, William J.
;     APPLICANT: Barrett, Ronald W.
;     APPLICANT: Cwirla, Steven E.
;     APPLICANT: Gates, Christian
;     APPLICANT: Schatz, Peter J.
;     APPLICANT: Balasubramanian, Palaniappan
;     APPLICANT: Wagstrom, Christopher R.
;     APPLICANT: Hendren, Richard W.
;     APPLICANT: Deprince, Randolph B.
;     APPLICANT: Podduturi, Surekha
;     APPLICANT: Yin, Qun
;   TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;   TITLE OF INVENTION: RECEPTOR
;   NUMBER OF SEQUENCES: 244
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Glaxo Wellcome
;     STREET: Five Moore Drive, P.O. Box 13398
;     CITY: Research Triangle Park
;     STATE: NC
;     COUNTRY: USA
;     ZIP: 27709
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/764,640
;       FILING DATE: 11-DEC-1996
;       CLASSIFICATION: 514
;       ADDRESSEE: Glaxo Wellcome
;       STREET: Five Moore Drive, P.O. Box 13398
;       CITY: Research Triangle Park
;       STATE: NC
;       COUNTRY: USA
;       ZIP: 27709
;   INFORMATION FOR SEQ ID NO: 185:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 15 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: linear
;       TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;     US-08-764-640-185
;
; Query Match          37.8%; Score 73; DB 2; Length 15;
; Best Local Similarity 100.0%; Pred. No. 0.0057;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLROWLAARA 14
;   |||||
; DB 2 IEGPTLROWLAARA 15
```

```

RESULT 14
US-08-973-225-17
; Sequence 17, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17

```

```

Query Match      37.8%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 IEGPTLROWLAARA 14
Db 1 IEGPTLROWLAARA 14

```

```

RESULT 15
US-08-973-225-185
; Sequence 185, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.

```

```

; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185

```

```

Query Match      37.8%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 IEGPTLROWLAARA 14
Db 2 IEGPTLROWLAARA 15

```

```

RESULT 16
US-09-244-298A-17
; Sequence 17, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

us-09-422-838c-30.ra1

Wed Oct 9 10:30:04 2002

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; MOLECULE TYPE: peptide
; US-09-244-298A-17

Query Match 37.8%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14
|||||

RESULT 17
US-09-244-298A-185
; Sequence 185, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-185

Query Match 37.8%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 2 IEGPTLRQWLAAARA 15
|||||

RESULT 18
US-09-516-704-17
; Sequence 17, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT:
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 37.8%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
|||||

```

```

Db      1  IEGPTLROWLAARA 14

RESULT 19
US-09-516-704-185
; Sequence 185, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
;            Barrett, Ronald W.
;            Cwirla, Steven E.
;            Gates, Christian
;            Schatz, Peter J.
;            Balasubramanian, Palaniappan
;            Wagstrom, Christopher R.
;            Hendren, Richard W.
;            Deprience, Randolph B.
;            Podduturi, Surekha
;            TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;            RECEPTOR
;            NUMBER OF SEQUENCES: 244
;            CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Glaxo Wellcome
;            STREET: Five Moore Drive, P.O. Box 13398
;            CITY: Research Triangle Park
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;            COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;            CURRENT APPLICATION DATA:
;            FILING DATE: 01-Mar-2000
;            APPLICATION NUMBER: US/09/516,704
;            CLASSIFICATION: <Unknown>
;            ATTORNEY/AGENT INFORMATION:
;            NAME: Hrubiec, Robert T.
;            REGISTRATION NUMBER: 36,392
;            REFERENCE/DOCKET NUMBER: PK3281
;            TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-248-1000
;            INFORMATION FOR SEQ ID NO: 185:
;            SEQUENCE CHARACTERISTICS:
;            LENGTH: 15 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            MOLECULE TYPE: peptide
;            SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-09-516-704-185
Query Match      37.8%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  IEGPTLROWLAARA 14
Db      2  IEGPTLROWLAARA 15

RESULT 20
US-08-764-640-18
; Sequence 18, Application US/08764640
; Patent No. 5869451
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
;            Barrett, Ronald W.
;            Cwirla, Steven E.
;            Gates, Christian
;            Schatz, Peter J.

```

```

; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
US-08-764-640-18
Query Match      37.8%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  IEGPTLROWLAARA 14
Db      1  IEGPTLROWLAARA 14

```

```

RESULT 21
US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244

```

us-09-422-838c-30.ra1

Wed Oct 9 10:30:04 2002

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/764,640
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-232
Query Match 37.8%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No: 0.0061;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
DB 2 IEPTLRQWLAARA 15

RESULT 23
US-08-973-225-18
Sequence 18, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide

```

```
;
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product="Beta-ala"
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18
    Query Match      37.8%; Score 73; DB 3; Length 16;
    Best Local Similarity 100.0%; Pred. No. 0.0061;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
    |||||
Db 1 IEGPTLQWLAAARA 14

RESULT 24
US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194
    Query Match      37.8%; Score 73; DB 3; Length 16;
    Best Local Similarity 100.0%; Pred. No. 0.0061;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
    |||||
Db 2 IEGPTLQWLAAARA 15
```

```
RESULT 25
US-08-973-225-220
; Sequence 220, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220
    Query Match      37.8%; Score 73; DB 3; Length 16;
    Best Local Similarity 100.0%; Pred. No. 0.0061;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
    |||||
Db 2 IEGPTLQWLAAARA 15

RESULT 26
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
```

```

; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
;
US-09-244-298A-18

```

```

Query Match 37.8%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 IEGPTLRQWLARA 14
| | | | | | | | | |
DB 1 IEGPTLRQWLARA 14

```

```

RESULT 27
US-09-244-298A-194
; Sequence 194, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC

```

```

; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-09-244-298A-194

```

```

Query Match 37.8%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 IEGPTLRQWLARA 14
| | | | | | | | | |
DB 2 IEGPTLRQWLARA 15

```

```

RESULT 28
US-09-244-298A-232
; Sequence 232, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392

```


us-09-422-838c-30.ra1

Wed Oct 9 10:30:04 2002

```

; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-232

Query Match 37.8%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

RESULT 29
US-09-516-704-18
; Sequence 18, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-09-516-704-194

Query Match 37.8%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

Search completed: October 9, 2002, 09:06:33
Job time : 5.98595 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8 09368 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838c-30
Perfect score: 193
Sequence: 1 IEQPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.5	32.4	209	2 B42687	neurotrophin-4 pre
2	61	31.6	500	2 T20961	hypothetical prote
3	60.5	31.3	488	2 G87033	probable ATP/grp-b
4	60.5	31.3	518	2 S72938	hflx protein - Myc
5	59.5	30.8	112	2 F70954	probable lsr2 prot
6	59.5	30.8	210	2 A42687	neurotrophin-4 pre
7	59	30.6	497	2 T35116	hypothetical prote
8	59	30.6	683	2 B71325	conserved hypothet
9	58.5	30.3	77	1 INSH	insulin precursor
10	58.5	30.3	105	1 IPBO	insulin precursor
11	58.5	30.3	495	2 D70505	probable HflX - My
12	58	30.1	339	2 S20880	homeotic protein H
13	56	29.0	103	2 T47718	hypothetical prote
14	56	29.0	201	2 T49792	hypothetical prote
15	56	29.0	249	2 T04436	ankyrin 3 homolog
16	56	29.0	303	2 S71185	splicing factor SF
17	56	29.0	331	2 T26807	hypothetical prote
18	56	29.0	333	2 T26808	hypothetical prote
19	56	29.0	465	2 S41644	polyadenylate-bind
20	56	29.0	591	1 F0VNMV	gag polyprotein -
21	56	29.0	619	1 KSNCLT	laccase (EC 1.10.3
22	56	29.0	619	1 KSNCLT	laccase (EC 1.10.3
23	56	29.0	3190	2 T13828	CREB-binding prote
24	55.5	28.8	328	2 G88499	protein K0467.10 (
25	55	28.5	201	2 J01094	hypothetical 20.2K
26	55	28.5	271	2 S34566	glycine-rich prote
27	55	28.5	327	2 AG1974	hypothetical prote
28	55	28.5	445	2 A49447	transcription fact
29	55	28.5	490	2 T09084	phosphatidylinosit

30	55	28.5	495	2 E70948	probable amidase -
31	55	28.5	767	2 E70895	hypothetical glyci
32	55	28.5	865	2 T34584	probable secreted
33	55	28.5	904	2 C70559	probable polA prot
34	54.5	28.2	112	2 B43601	LSR2 T-cell antige
35	54.5	28.2	209	2 C97617	glutathione S-tran
36	54.5	28.2	226	2 A75288	hypothetical prote
37	54.5	28.2	562	2 F72771	hypothetical prote
38	54	28.0	177	2 C88115	probable lysyl-PRN
39	54	28.0	180	2 T49530	protein F53C3.3 [1
40	54	28.0	309	2 T19389	related to glycine
41	54	28.0	497	2 T45406	hypothetical prote
42	54	28.0	545	2 D87259	probable amidase (
43	54	28.0	556	2 A32466	phosphoglucomutase
44	54	28.0	702	2 G01840	numb protein - fru
45	53.5	27.7	198	2 A54507	T-box protein 2 -
					dnak-type molecula

ALIGNMENTS

RESULT 1
B42687
neurotrophin-4 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
C:Accession: B42687; JH0504; JH0505
R:IP, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Balluscio, Proc. Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992
A:Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distri
A:Reference number: A42687; MUID:92212967
A:Accession: B42687
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <IPA>
A:Cross-references: GB:M66742; NID:g205775; PIDN:AAA41728.1; PID:g205776
R:Berkemeyer, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosenth Neuron 7, 857-866, 1991
A:Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.
A:Reference number: JH0503; MUID:92075279
A:Accession: JH0504
A:Molecule type: DNA
A:Residues: 1-209 <BER>
A:Accession: JH0505
A:Molecule type: mRNA
A:Residues: 1-176, 'P', 178-209 <BER1>
A:Cross-references: GB:S69323; NID:g240025; PIDN:AAB20548.1; PID:g240026
C:Comment: This protein is a targeted-derived, diffusible neurotrophic factor.
C:Comment: The neurotrophins stimulate autophosphorylation and transduce signals thro
C:Superfamily: nerve growth factor beta chain
C:Keywords: glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-79/Domain: propeptide #status predicted <PRO>
F:80-209/Product: neurotrophin-5 #status predicted <NEU>
F:75/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 32.4%; Score 62.5; DB 2; Length 209;
Best Local Similarity 37.5%; Pred. No. 5.7;
Matches 15; Conservative 2; Mismatches 14; Indels 9; Gaps 1;

QY 3 GPTLRQWL-----AARAGGGGGGIEGPTLRQWL 33
| | | | |
Db 128 GSPLOYFFETRCKAESAGEGGVGGGCRGVDRHRLS 167

RESULT 2
T20961
hypothetical protein F1589.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20961
R:Percy, C.
submitted to the EMBL Data Library, August 1996

us-09-422-838c-30.rpr

Wed Oct 9 10:30:05 2002

Best Local Similarity 43.3%; Pred. No. 22; Mismatches 13; Conservative 2; Indels 8; Gaps 1;

QY 4 PTLROW-----LAARAGGKGGGIEGP 26
 Db 219 PRLRGWESMSRQVGGGSGGGVGLRGP 248

RESULT 5

F70954
 Probable lsr2 protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: F70954
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, C.; Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Rajandream, M.A.; 537-544, 1998
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70954
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-112 <COL>
 A:Cross-references: GB:295557; GB:AL123456; NID:g3242276; PIDN:CAB08947.1; PID:g21139
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: lsr2

Query Match

30.8%; Score 59.5; DB 2; Length 112;

Best Local Similarity

33.3%; Pred. No. 6.9; Mismatches 7; Indels 13; Gaps 2;

Matches

13; Conservative

QY

6 LQWLAA-----RAGGKGGG---EPTLRQW 31

Db 48 LQWVAGRRVGGRRGRSGRGAIDREQSAIREW 86

RESULT 6

A42687

Neurotrophin-4 precursor - human

N:Alternate names: neurotrophin-5

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999

C:Accession: A42687; JH0503

R: Ip, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Belluscio, R.; Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992

A:Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distribution

A:Reference number: A42687; MUID:92212967

A:Accession: A42687

A:Molecule type: DNA

A:Residues: 1-210 <IPI>

A:Cross-references: GB:M86528; NID:g190264; PIDN:AAA60154.1; PID:g190265

A:Note: sequence extracted from NCBI backbone (NCBI:93810, NCBI:93811)

R: Berkemeier, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosenthal, N. 7, 857-866, 1991

A:Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.

A:Reference number: JH0503; MUID:92075279

A:Accession: JH0503

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-210 <BER>

C:Comment: The neurotrophins stimulate autophosphorylation and transduce signals through

C:Genetics:
 C:Gene: GDB:NTF5
 A:Cross-references: GDB:134723; OMIM:162662
 A:Map position: 19pter-19qter
 C:Superfamily: nerve growth factor beta chain
 C:Keywords: glycoprotein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-80/Domain: propeptide #status predicted <PRO>
 F:81-210/Product: neurotrophin-4 #status predicted <NEU>

Query Match

31.3%; Score 60.5; DB 2; Length 518;

Best Local Similarity

43.3%; Pred. No. 20; Mismatches 8; Indels 7; Gaps 1;

Matches

13; Conservative

QY

4 PTLROW-----LAARAGGKGGGIEGP 26

Db 189 PRLRGWESMSRQVGGGSGGGVGLRGP 218

RESULT 4

S72938

hflX protein - Mycobacterium leprae

N:Alternate names: B2235_C2_202 protein

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72938

R: Smith, D.R.; Robinson, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B2235.

A:Reference number: S72587

A:Accession: S72938

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-518 <SMI>

A:Cross-references: EMBL:U00019; NID:g467079; PIDN:AAAI7274.1; PID:g467091

C:Genetics:
 C:Start codon: GTG
 C:Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match

31.3%; Score 60.5; DB 2; Length 518;

Best Local Similarity

43.3%; Pred. No. 20; Mismatches 8; Indels 7; Gaps 1;

Matches

13; Conservative

QY

4 PTLROW-----LAARAGGKGGGIEGP 26

Db 189 PRLRGWESMSRQVGGGSGGGVGLRGP 218

F:76/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 30.8% Score 59.5; DB 2; Length 210;
Best Local Similarity 35.0%; Pred. No. 12;
Matches 14; Conservative 3; Mismatches 14; Indels 9; Gaps 1;

QY 3 GPTLROWL-----AARAGGKGGGIEGPTLROWLA 33
DB 129 GSPURQYFFETRCADNAEEGPGAGGCGRGVDRHHWS 168

RESULT 7
T35116
hypothetical protein SC4H2.17 SC4H2.17 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 02-Sep-2000
R:Accession: T35116
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z21568
A:Accession: T35116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-497 <SEE>
A:Cross-references: EMBL:AL022268; PIDN:CAA18333.1; GSPDB:GN00070; SCOEDB:SC4H2.17
A:Experimental source: strain A3(2)
C:Genetics:
C:Gene: SCOEDB:SC4H2.17
C:Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 30.6% Score 59; DB 2; Length 497;
Best Local Similarity 51.9%; Pred. No. 30;
Matches 14; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

QY 4 PTLROW---LAARAGGKGGG-GIEGP 26
DB 206 PRLRGWQSLRQMGKGKGGGLATRGP 232

RESULT 8
B71325
conserved hypothetical protein TP0421 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
R:Accession: B71325
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn
Lyon, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: B71325
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-683 <COL>
A:Cross-references: GB:AE001220; GB:AE000520; MID:g3322705; PIDN:AAC65409.1; PID:g332270
A:Experimental source: strain Nichols
C:Genetics:
C:Gene: TP0421

Query Match 30.6% Score 59; DB 2; Length 683;
Best Local Similarity 43.8%; Pred. No. 40;
Matches 14; Conservative 2; Mismatches 12; Indels 4; Gaps 1;

QY 4 PTLROWLAARAGGKGGGIEGPTLROWLAAR 35
DB 74 PLILEWL-----GNAYYRGIEGALHOWGAAR 101

RESULT 9
INSH
insulin precursor - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: S16430; S16431
R:Brown, H.; Sanger, F.; Kitai, R.
Biochem. J. 60, 556-565, 1955
A:Title: The structure of pig and sheep insulins.
A:Reference number: A90344
A:Accession: S16430
A:Molecule type: protein
A:Residues: 1-30; 57-77 <BRO>
R:Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin
A:Reference number: A92111; MUID:72258016
A:Accession: S16431
A:Molecule type: protein
A:Residues: 31-56 <PET>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:31-56/Domain: connecting peptide #status experimental <MAT>
F:57-77/Domain: insulin chain A #status experimental <CPEP>
F:7-63, 19-76, 82-67/Disulfide bonds: #status predicted

Query Match 30.3% Score 58.5; DB 1; Length 77;
Best Local Similarity 50.0%; Pred. No. 6.3;
Matches 13; Conservative 2; Mismatches 8; Indels 3; Gaps 1;

QY 1 IEPTLROWLAARAGGKGGGIEGP 26
DB 32 VEGP---QVGALELAGGPGAGGIEGP 54

RESULT 10
IPBO
insulin precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence_revision 22-Apr-1995 #text_change 16-Jul-1999
C:Accession: A40909; A92080; A91185; A90342; A90341; S48184; S48185; S46258;
R:D'Agostino, J.; Younes, M.A.; White, J.W.; Besch, P.K.; Field, J.B.; Frazier, M.L.
Mol. Endocrinol. 1, 327-331, 1987
A:Title: Cloning and nucleotide sequence analysis of complementary deoxyribonucleic a
A:Reference number: A40909; MUID:88288209
A:Accession: A40909
A:Molecule type: mRNA
A:Residues: 1-105 <DAA>
A:Cross-references: GB:M54979; MID:g163578; PIDN:AAA30722.1; PID:g163579
A:Experimental source: fetal pancreas
R:Nolan, C.; Margolish, E.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 2780-2795, 1971
A:Title: The structure of bovine proinsulin.
A:Reference number: A92080; MUID:71166442
A:Accession: A92080
A:Molecule type: protein
A:Residues: 25-105 <NOL>
R:Steiner, D.F.; Cho, S.; Oyer, P.E.; Terris, S.; Peterson, J.D.; Rubenstein, A.H.
J. Biol. Chem. 246, 1365-1374, 1971
A:Title: Isolation and characterization of proinsulin C-peptide from bovine pancreas.
A:Reference number: A92074; MUID:71116409
A:Accession: A92074
A:Molecule type: protein
A:Residues: 57-82 <STE>
R:Saikangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 183-189, 1971
A:Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancre
A:Reference number: A91185; MUID:71257721
A:Accession: A91185
A:Molecule type: protein
A:Residues: 57-82 <SAL>
R:Sanger, F.; Thompson, E.O.P.
Biochem. J. 53, 366-374, 1953
A:Title: The amino-acid sequence in the glycyl chain of insulin. 2. The investigation
A:Reference number: A90342

Wed Oct 9 10:30:05 2002

C:Superfamily: GTP-binding protein hflx: translation elongation factor Tu homology

Query Match 30.3%; Score 58.5; DB 2; Length 495;

Best Local Similarity 43.3%; Pred. No. 34;
Matches 13; Conservative 1; Mismatches 9; Indels 7; Gaps 1;

QY 4 PTLRQWLARAGGKGGGIEGP 26

DB 199 PRLRGESMSRQAGRGAGGGGGLRGP 228

RESULT 12

S20880

homeotic protein Hox 4.5 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 17-Nov-2000

C:Accession: S20880; S09569; S09398

R:Renucci, A.; Zappavigna, V.; Zakany, J.; Izpisua-Belmonte, J.C.; Buerki, K.; Duboul

EMBO J. 11, 1459-1468, 1992

A:Title: Comparison of mouse and human HOX-4 complexes defines conserved sequences in

A:Reference number: S20879; MUID:92224884

A:Accession: S20880

A:Molecule type: DNA

A:Residues: 1-339 <REN>

A:Cross-references: EMBL:X62669; NID:g51414; PIDN:CAA44542.1; PID:g51416

R:Duboule, D.; Dolle, P.

EMBO J. 8, 1497-1505, 1989

A:Title: The structural and functional organization of the murine HOX gene family res

A:Reference number: S09569; MUID:89356621

A:Accession: S09569

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 272-331 <DUB>

A:Cross-references: EMBL:X14714; NID:g51427; PIDN:CAB57813.1; PID:g6015583

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1989

R:Dolle, P.; Duboule, D.

EMBO J. 8, 1507-1515, 1989

A:Title: Two gene members of the murine HOX-5 complex show regional and cell-type spe

A:Reference number: S09398; MUID:89356622

A:Accession: S09398

A:Molecule type: DNA

A:Residues: 272-331 <DOL>

A:Cross-references: GB:X14714; GB:M21040; NID:g51427; PIDN:CAB57813.1; PID:g6015583

C:Genetics:

A:Gene: Hox-4.5

A:Introns: 260/1

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:273-329/Domain: homeobox homology <HOX>

Query Match 30.1%; Score 58; DB 2; Length 339;

Best Local Similarity 40.6%; Pred. No. 27;

Matches 13; Conservative 2; Mismatches 9; Indels 8; Gaps 1;

QY 3 GPTLRQWLARAGGKGGGIEGP 26

DB 101 GRVVRSMWEPLPGFPGAGGGGGGGGGP 132

RESULT 13

T47718

hypothetical protein FL116.200 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47718

R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer,

submitted to the Protein Sequence Database, March 2000

A:Reference number: T24473

A:Accession: T47718

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-103 <BEN>

A:Cross-references: EMBL:AL161667

A:Accession: A90342

A:Molecule type: protein

A:Residues: 85-105 <SANG>

R:Sanger, F.; Tuppy, H.

Biochem. J. 49, 481-490, 1951

A:Title: The amino-acid sequence in the phenylalanyl chain of insulin..2. The investigat

A:Reference number: A90341

A:Accession: A90341

A:Molecule type: protein

A:Residues: 25-54 <SA2>

R:Cheng, R.; Kawakishi, S.

Eur. J. Biochem. 223, 759-764, 1994

A:Title: Site-specific oxidation of histidine residues in glycosylated insulin mediated by O

A:Reference number: S48184; MUID:94333378

A:Accession: S48184

A:Molecule type: protein

A:Residues: 85-105 <CHE>

A:Accession: S48185

A:Status: preliminary

A:Molecule type: protein

A:Residues: 25-30, 'X', 32-42, 'X', 44-54 <CH2>

R:Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.

Biochem. J. 60, 541-556, 1955

A:Title: The disulphide bonds of insulin.

A:Reference number: A90343

A:Contents: annotation; amides; disulfides

R:Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W.

FEBS Lett. 349, 205-209, 1994

A:Title: Existence of a molecular ruler in proteasomes suggested by analysis of degradat

A:Reference number: S46258; MUID:94326921

A:Accession: S46258

A:Status: preliminary

A:Molecule type: protein

A:Residues: 25-54 <WEN>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-54/Domain: insulin chain B #status experimental <BCH>

F:55-82/Domain: connecting peptide #status experimental <CPPE>

F:83-105/Domain: insulin chain A #status experimental <ACH>

F:31-91,43-104,90-95/Disulfide bonds: #status experimental

Query Match 30.3%; Score 58.5; DB 1; Length 105;

Best Local Similarity 50.0%; Pred. No. 8.3;

Matches 13; Conservative 2; Mismatches 8; Indels 3; Gaps 1;

QY 1 IEGPTLRQWLARAGGKGGGIEGP 26

DB 58 VEGP---QVGALELAGGPGAGGLEGP 80

RESULT 11

D70505

probable Hflx - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 02-Sep-2000

C:Accession: D70505

R:Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

R:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: D70505

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-495 <COL>

A:Cross-references: GB:Z98209; GB:AL123456; NID:g3261838; PIDN:CAB10901.1; PID:e332282;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: hflx

S71185
splicing factor SF-2 homolog - *Arabidopsis thaliana*

C; Genetics:
A; Gene: CESP:Y41C4A.4b
A; Introns: 24/3; 50/2; 81/3; 161/1; 230/1; 294/3
C; Superfamily: fos/jun DNA-binding domain homolog

Wed Oct 9 10:30:05 2002

us-09-422-838c-30.rpr

Query Match 29.0%; Score 56; DB 2; Length 333;
 Best Local Similarity 69.2%; Pred. No. 44;
 Matches 9; Conservative 2; Mismatches 0; Gaps 0;

QY 15 GCGKGGGGIEGPT 27

Db 169 GGGGGGGVGP 181

RESULT 19

S41644 Polyadenylate-binding protein - fruit fly (Drosophila melanogaster)

N:Alternate names: rox8 protein

C:Species: Drosophila melanogaster

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999

C:Accession: S41644

R:Brand, S.; Bourbon, H.M.

Nucleic Acids Res. 21, 3699-3704, 1993

A:Title: The developmentally-regulated Drosophila gene rox8 encodes an RRM-type RNA binding protein

A:Reference number: S41644; MUID:93376491

A:Accession: S41644

A:Molecule type: DNA

A:Residues: 1-465

A:Cross-references: EMBL:L13037; NID:g304797; PIDN:AAA28828.1; PID:g304798

A:Accession: S41645

A:Molecule type: mRNA

A:Residues: 1-465

A:Cross-references: EMBL:L13038; NID:g304799; PIDN:AAA02941.1; PID:g304800

C:Genetics:

A:Gene: FlyBase:Rox8

A:Cross-references: FlyBase:FBgn0005649

A:Introns: 76/3; 379/3; 411/1

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

C:Keywords: RNA binding

F:8-72/Domain: ribonucleoprotein repeat homology <RRM1>

F:139-282/Domain: ribonucleoprotein repeat homology <RRM2>

F:144-480/Domain: ribonucleoprotein repeat homology <RRM3>

F:222-284/Domain: ribonucleoprotein repeat homology <RRM3>

Query Match 29.0%; Score 56; DB 2; Length 465;

Best Local Similarity 34.2%; Pred. No. 60;

Matches 13; Conservative 4; Mismatches 3; Indels 18; Gaps 1;

QY 7 ROWLAARA-----GGKGGGGIEGPT 26

Db 158 ROWTASRSITNWSTRKLPPTPPKSGGGGGGGGGP 195

RESULT 20

FOMVNM

gag polyprotein - mouse mammary tumor virus

C:Species: mouse mammary tumor virus, MMTV

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999

C:Accession: A26795

R:Moore, R.; Dixon, M.; Smith, R.; Peters, G.; Dickson, C.

J. Virol. 61, 480-490, 1987

A:Title: Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus:

A:Reference number: A93030; MUID:87112944

A:Accession: A26795

A:Molecule type: DNA

A:Residues: 1-591 <MOO>

A:Cross-references: EMBL:M15122; NID:g332127; PIDN:AAA46543.1; PID:g332130

C:Genetics:

A:Gene: gag

C:Superfamily: mouse mammary tumor virus gag polyprotein

C:Keywords: core protein; DNA binding; polyprotein

Query Match 29.0%; Score 56; DB 1; Length 591;

Best Local Similarity 83.3%; Pred. No. 74;

Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 15 GCGKGGGGIEGPT 26

Db 514 GCGKGGGGIEGPT 525

RESULT 21

KSNCLT

laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)

N:Alternate names: urishiol oxidase

C:Species: Neurospora crassa

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999

C:Accession: A28523; A29762

R:Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.

J. Biol. Chem. 263, 885-896, 1988

A:Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino-

A:Reference number: A28523; MUID:88087214

A:Accession: A28523

A:Molecule type: DNA

A:Residues: 1-619 <GER>

A:Cross-references: EMBL:M14554

R:Germann, U.A.; Lerch, K.

Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986

A:Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospor

A:Reference number: A29762; MUID:87067412

A:Accession: A29762

A:Molecule type: DNA

A:Residues: 379-619 <GE2>

A:Cross-references: GB:M14554; NID:g168823; PIDN:AAA33590.1; PID:g168824

C:Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquino

C:Genetics:

C:Superfamily: laccase

C:Keywords: copper; glycoprotein; oxidoreductase

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-49/Domain: propeptide #status predicted <PRO>

F:50-619/Domain: laccase #status predicted <MAT>

F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>

F:216-372/Domain: middle beta-barrel #status predicted <BB2>

F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>

F:139-282/Domain: binding site: carboxylate (Asn) (covalent) #status predicted

F:144-480/Binding site: copper (His) (type 2) #status predicted

F:146-189, 191, 482, 548, 550/Binding site: 2Cu-O cluster (His) (copper type 3) #status p

F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 29.0%; Score 56; DB 1; Length 619;

Best Local Similarity 63.6%; Pred. No. 78;

Matches 14; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 11 AARAGGGGGGIEGPTLRQ-W 31

Db 44 AERYGGG-GGGCNSPTNRQW 64

RESULT 22

KSNCLT

laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)

N:Alternate names: urishiol oxidase

C:Species: Neurospora crassa

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999

C:Accession: B28523

R:Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.

J. Biol. Chem. 263, 885-896, 1988

A:Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino-

A:Reference number: A28523; MUID:88087214

A:Accession: B28523

A:Molecule type: DNA

A:Residues: 1-619 <GER>

A:Cross-references: EMBL:M18334; NID:g168827; PIDN:AAA33592.1; PID:g168828

C:Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquino

C:Genetics:

A:Introns: 86/3

C:Superfamily: laccase

C:Keywords: copper; glycoprotein; oxidoreductase

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-49/Domain: propeptide #status predicted <PRO>

F:50-619/Domain: laccase #status predicted <MAT>

F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
F:216-372/Domain: middle beta-barrel #status predicted <BB1>
F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB2>
F:139, 282, 295, 340, 422, 444/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:144, 480/Binding site: copper (His) (type 2) #status predicted
F:146, 189, 191, 482, 548, 550/Binding site: 2Cu-O cluster (His) (copper type 3) #status predicted
F:477, 549, 554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 29.0%; Score 56; DB 1; Length 619;
Best Local Similarity 63.6%; Pred No. 78;
Matches 14: Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Oy 11 AARAGGGKGGGIEGPTLRQ-W 31
| | | | | | | | | |
Db 44 AERYGGG-GGGGCNSPTNRQCW 64

RESULT 23
Tl3828
CREB-binding protein homolog - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: Tl3828
R:Akimaru, H.; Chen, Y.; Dai, P.; Hou, D.X.; Nonaka, M.; Smolik, S.M.; Armstrong, S.; G
Nature 386, 735-738, 1997
A>Title: *Drosophila* CBP is a co-activator of cubitus interruptus in hedgehog signalling
A:Reference number: Z17785; MUID:97263578
A:Accession: Tl3828
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-3190 <AKI>
A:Cross-references: EMBL:U88570; NID:g1916929; PID:g1916930; PIDN:AAB53050.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0015624
A:Map position: X
C:Superfamily: bromodomain homology
F:1723-1780/Domain: bromodomain homology <BRO>

Query Match 29.0%; Score 56; DB 2; Length 3190;
Best Local Similarity 61.1%; Pred No. 3.4e+02;
Matches 11: Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 10 LAARAGGGKGGGIEGPT 27
| | | | | | | | | |
Db 44 LTGGAGGGNGGGASGV 61

RESULT 24
G88499
protein K04G7.10 [imported] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: G88499
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88499
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <STO>
A:Cross-references: GB:chr_III; PIDN:AA62536.1; PID:g687845; GSPDB:GN00021; CESP:K04G7.
C:Genetics:
A:Gene: K04G7.10
A:Map position: 3

Query Match 28.8%; Score 55.5; DB 2; Length 328;
Best Local Similarity 32.6%; Pred No. 49;
Matches 14: Conservative 1; Mismatches 7; Indels 21; Gaps 1;

Oy 3 GPTLRQLAARAGGGKG-----GGGIE 24

us-09-422-838c-30.rpr

Wed Oct 9 10:30:05 2002

Db 231 GP-----LLAAGGGGGGGSGSGSGTARDEWL 260

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr1346

Query Match 28.5%; Score 55; DB 2; Length 327;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 11; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 10 LAARAGGKGKGIEGPTLRQW 31

Db 36 LAARSGGRIGGGFRAPSSRTY 57

RESULT 28

A49447 transcription factor Brn-2 - rat

N:Alternate names: class III POU domain protein brain-2

C:Species: Rattus norvegicus (Norway rat)

C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 20-Feb-1998

A:Accession: A49447

R:Li, P.; He, X.; Guerrero, M.R.; Mok, M.; Aggarwal, A.; Rosenfeld, M.G.

Genes Dev. 7, 2483-2496, 1993

A:Title: Spacing and orientation of bipartite DNA-binding motifs as potential functional

A:Reference number: A49447; MUID:94102531

A:Accession: A49447

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-445

A:Cross-references: GB:L27663; NID:9443687

A:Experimental source: brain

A:Note: sequence extracted from NCBI backbone (NCBIP:141696)

C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:68-90/Region: glycine-rich

F:125-151/Region: glutamine-rich

F:153-165/Region: histidine/proline-rich

F:213-261/Region: histidine/proline-rich

F:271-338/Domain: POU domain homology <POU>

F:357-413/Domain: homeobox homology <HOX>

Query Match 28.5%; Score 55; DB 1; Length 445;

Best Local Similarity 60.0%; Pred. No. 74;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 8 QWLAARAGGKGKG 22

Db 60 QWITALSHGSGGGG 74

RESULT 29

T09084 phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)

C:Species: Chlamydomonas reinhardtii

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

A:Accession: T09084

R:Molendijk, A.J.; Irvine, R.F.

Plant Mol. Biol. 37, 53-66, 1998

A:Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinositol

A:Reference number: Z16411; MUID:98281574

A:Accession: T09084

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-490 <MOL>

A:Cross-references: EMBL:U97663; NID:g2109290; PIDN:AA50018.1; PID:g2109291

A:Experimental source: strain CW-15

C:Genetics:

A:Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match 28.5%; Score 55; DB 2; Length 490;

Best Local Similarity 45.7%; Pred. No. 81;

Matches 16; Conservative 2; Mismatches 7; Indels 10; Gaps 3;

Qy 3 GPTLRQWLAARAGGKGKG 32

RESULT 30

E70948 probable amidase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

A:Accession: E70948

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: E70948

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-495 <COL>

A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16640.1; PID:g282

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3175

C:Superfamily: indoleacetamide hydrolase

Query Match 28.5%; Score 55; DB 2; Length 495;

Best Local Similarity 55.0%; Pred. No. 81;

Matches 11; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 GPTLRQWLAARAGGKGKG 22

Db 147 GRTNPNWDAARTSGGSAGG 166

Search completed: October 9, 2002, 09:05:09

Job time: 9.09368 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-30

Perfect score: 193
Sequence: 1 IEPTLRQWLAAAGGGGGIEGTTLRQWLAARA 36

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.5	32.4	209	1 NT5_RAT	P34131 rattus norv
2	59.5	30.8	112	1 LSR2_MYCTU	O06285 mycobacteri
3	59.5	30.8	210	1 NT5_HUMAN	P34130 homo sapien
4	59	30.6	497	1 FXD2_HUMAN	O60548 homo sapien
5	58.5	30.3	105	1 INS_BOVIN	P01317 bos taurus
6	58.5	30.3	105	1 INS_SHEEP	P01318 ovis aries
7	58	30.1	339	1 HXD9_MOUSE	P28357 mus musculu
8	57.5	29.8	266	1 SCO2_HUMAN	O43819 homo sapien
9	57	29.5	694	1 FZD8_HUMAN	P09461 homo sapien
10	56	29.0	591	1 GAG_MTVB	P10258 mouse mamma
11	56	29.0	619	1 LAC1_NEUCR	P06811 neurospora
12	56	29.0	619	1 LAC2_NEUCR	P10574 neurospora
13	55	28.5	201	1 YR21_TRSVR	P25245 tomato ring
14	55	28.5	904	1 DP01_MYCTU	O07700 mycobacteri
15	54.5	28.2	112	1 LSR2_MYCLE	P24094 mycobacteri
16	54.5	28.2	562	1 SYK_AERPE	O9yft9 aeropyrum p
17	54	28.0	497	1 GATA_MYCLE	O33105 mycobacteri
18	54	28.0	556	1 NUMB_DROME	P16554 drosophila
19	54	28.0	702	1 TBX2_RAT	Q13207 homo sapien
20	54	28.0	716	1 E2BE_RAT	Q64350 rattus norv
21	54	28.0	1048	1 AG01_ARATH	O04379 arabidopsis
22	53.5	27.7	198	1 HS70_SCHJA	P12795 schistosoma
23	53.5	27.7	969	1 PAC4_HUMAN	P29122 homo sapien
24	53	27.5	332	1 YACO_ALCEU	P31640 alcaligenes
25	53	27.5	333	1 SIX3_MOUSE	O62233 mus musculu
26	53	27.5	394	1 FXD3_CHICK	P79772 gallus gall
27	53	27.5	426	1 HKLB_LYCES	O22300 lycopersico
28	53	27.5	443	1 OC3N_HUMAN	P20265 homo sapien
29	53	27.5	445	1 OC3N_MOUSE	P31360 mus musculu
30	53	27.5	494	1 GATA_MYCTU	O53258 mycobacteri
31	53	27.5	504	1 ATIN_HSVBP	P30020 bovine heip
32	53	27.5	584	1 CNAL_DROME	P12252 drosophila
33	53	27.5	593	1 K1CJ_HUMAN	P13645 homo sapien

34	53	27.5	1001	1	ORK1_DROME	Q94526 drosophila
35	53	27.5	1168	1	MYSC_ACACA	P10569 acanthamoeb
36	53	27.5	1178	1	PHYB_SORBI	P93527 sorghum bic
37	53	27.5	2142	1	PAT2_HUMAN	P48634 homo sapien
38	52.5	27.2	174	1	SSB_HHOSH	Q924q8 rhodobacter
39	52.5	27.2	864	1	KLTK_HUMAN	P29376 homo sapien
40	52	26.9	323	1	JUND_CHICK	P27921 gallus gall
41	52	26.9	348	1	SXL_CERCA	O61374 ceratitis c
42	52	26.9	440	1	DCO_DROME	O76324 drosophila
43	52	26.9	475	1	EVX2_MOUSE	P43749 mus musculu
44	52	26.9	569	1	K1CJ_MOUSE	P02535 mus musculu
45	52	26.9	888	1	KLTK_MOUSE	P08923 mus musculu

ALIGNMENTS

RESULT 1
NT5_RAT
ID NT5_RAT STANDARD; PRT; 209 AA.
AC P34131;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-5 precursor (NT-5) (Neutrophic factor 5) (Neurotrophin-4)
DE (NT-4) (Neutrophic factor 4).
GN NTF5 OR NTF4 OR NT4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92212967; PubMed=1313578;
RA Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
RA Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
RA Yancopoulos G.D.;
RT "Mammalian neurotrophin-4: structure, chromosomal localization,
RT tissue distribution, and receptor specificity";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92075279; PubMed=1742028;
RA Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
RA Rosenthal A.;
RT "Neurotrophin-5: a novel neurotrophic factor that activates trk and
RT trkB";
RL Neuron 7:857-866(1991).
CC -!- FUNCTION: COULD SERVE AS A TARGET-DERIVED TROPHIC FACTOR FOR
CC SENSORY AND SYMPATHETIC NEURONS
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, MUSCLE, OVARY, BRAIN,
CC HEART, STOMACH AND KIDNEY. EXPRESSED IN BOTH EMBRYO AND ADULT
CC TISSUES.
CC -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M86742; AAA41728.1; -;
CC EMBL; S69323; AAB20548.1; -;
CC PIR; JH0504; JH0504.
CC PIR; B42687; B42687.
CC HSSP; P34130; I88M.
CC InterPro; IPR002072; NGF.
CC Pfam; PF00243; NGF. 1.
CC PRINTS; PR00268; NGF. 1.
CC ProDom; PD002052; NGF. 1.
CC SMART; SM00140; NGF. 1.

Wed Oct 9 10:30:06 2002

```

DR PROSITE: PS00248; NGF_1; 1.
DR PROSITE: PS00270; NGF_2; 1.
KW Growth factor; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 79 NEUTROTROPHIN-5.
FT PROPEP 80 209 BY SIMILARITY.
FT CHAIN 80 209 BY SIMILARITY.
FT DISULFID 96 169 BY SIMILARITY.
FT DISULFID 140 198 BY SIMILARITY.
FT DISULFID 157 200 BY SIMILARITY.
FT CARBOHYD 75 75 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 177 177 R -> P (IN REF. 2).
SQ SEQUENCE 209 AA; 22332 MW; DF5112C05C5D5B85 CRC64;

Query Match 32.4%; Score 62.5; DB 1; Length 209;
Best Local Similarity 37.5%; Pred. No. 3.3;
Matches 15; Conservative 2; Mismatches 14; Indels 9; Gaps 1;

QY 3 GPTLRQWL-----AARAGGKGGGGEGPTLRQWLA 33
| | | | | | | | | | | | | | | | | | | |
Db 128 GSPLRQYFFETCKAESAGSGGPGVGGGGRGVDRRHWS 167

RESULT 2
LSR2_MYCTU STANDARD; PRT; 112 AA.
ID LSR2_MYCTU
AC O06285;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE LSR2 protein precursor.
DE LSR2 OR RV3597C OR MT3704 OR MTCV07H7B.25.
GN Mycobacterium tuberculosis.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K.K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisbal W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DOMINANT T-CELL ANTIGEN AND STIMULATES
LYMPHOPROLIFERATION (BY SIMILARITY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; 295557; CAB08947.1;

DR EMBL; AE007170; AAK48061.1; -.
DR TIGR; MT3704; -.
DR TubercuList; RV3597c; -.
KW Antigen; Immune response; Signal; Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 112 LSR2 PROTEIN.
SQ SEQUENCE 112 AA; 12098 MW; A4B32E478CBAC3E4 CRC64;

Query Match 30.8%; Score 59.5; DB 1; Length 112;
Best Local Similarity 33.3%; Pred. No. 3.9;
Matches 13; Conservative 7; Mismatches 6; Indels 13; Gaps 2;

QY 6 LKQWLA-----RAGGKGGGGI---EGPTLRQW 31
| | | | | | | | | | | | | | | | | | | |
Db 48 LKQWVAAGRRVGGRRGRSGSGRGAIDREQSAAIREW 86

RESULT 3
NTS_HUMAN STANDARD; PRT; 210 AA.
ID NTS_HUMAN
AC P34130;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-5 precursor (NT-5) (Neurotrophic factor 5) (Neurotrophin-4)
DE (NT-4) (Neurotrophic factor 4).
DE NTF5 OR NTF4.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=92212967; PubMed=1313578;
RA Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
Yancopoulos G.D.;
RA "Mammalian neurotrophin-4: structure, chromosomal localization,
tissue distribution, and receptor specificity.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92075279; PubMed=1742028;
RA Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
Rosenthal A.;
RA "Neurotrophin-5: a novel neurotrophic factor that activates trk and
trkB.";
RL Neuron 7:857-866(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
RX MEDLINE=20095835; PubMed=10631974;
RA Robinson R.C., Radziejewski C., Spraggon G., Greenwald J.,
Kostura M.R., Burtnick L.D., Stuart D.I., Choe S., Jones E.Y.;
RA "The structures of the neurotrophin 4 homodimer and the brain-derived
neurotrophic factor/neurotrophin 4 heterodimer reveal a common Trk-
binding site.";
RL Protein Sci. 8:2589-2597(1999).
CC -1- FUNCTION: TARGET-DERIVED SURVIVAL FACTOR FOR PERIPHERAL SENSORY
SYMPATHETIC NEURONS.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN PROSTATE, LOWER LEVELS
IN RHYMUS, PLACENTA, AND SKELETAL MUSCLE. EXPRESSED IN EMBRYONIC
AND ADULT TISSUES.
CC -1- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```

DR EMBL; M86528; AAA60154.1; -.
DR PIR; JH0503; JH0503.
DR PIR; A42687; A42687.
DR PDB; 1B8M; 09-FEB-99.
DR PDB; 1B98; 26-FEB-99.
DR MIM; 162662; -.
DR InterPro; IPR002072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PR00268; NGF.
DR PRODOM; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF; 1; 1.
DR PROSITE; PS00270; NGF; 2; 1.
DR Growth factor; Signal; 3D-structure.
KW SIGNAL 1 24
FT PROPEP 25 80
FT CHAIN 81 210 NEUROTROPHIN-5.
FT DISULFID 97 170
FT DISULFID 141 199
FT DISULFID 158 201
FT CARBOHYD 76 76
SQ SEQUENCE 210 AA; 22426 MW; DBC6A30195E139AD CRC64;
Query Match 30.8%; Score 59.5; DB 1; Length 210;
Best Local Similarity 35.0%; Pred. No. 6.9;
Matches 14; Conservative 3; Mismatches 14; Indels 9; Gaps 1;
QY 3 GPTLRQWL-----AARAGGGKGGGIEGPTLRQWLA 33
| | | | | | | | | | | | | | | | | |
DB 129 GSPLRQYFFETRCKADNAEGGPGAGGGCGVDRRHWS 168
| | | | | | | | | | | | | | | | | |

RESULT 4
FXD2_HUMAN STANDARD; PRT; 497 AA.
AC O60548;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-
related transcription factor 9) (FREAC-9).
GN FOXD2 OR FKHL17 OR FREAC9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98066765; PubMed=9403061;
RA Ernstsson S., Betz R., Lagercrantz S., Larsson C., Erickson S.,
RA Cederberg A., Carlsson P., Enerbaeck S.;
RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
expressed human forkhead gene that maps to chromosome 1p32-p34.";
RL Genomics 46:78-85(1997).
RN [2]
RP REVISIONS.
RA Enerbaeck S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF042832; AAC15421.1; -.
DR HSP; Q63245; 2HPF.

```

```

DR TRANSFAC; T02485; -.
DR MIM; 602211; -.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 90 94 POLY-ALA.
FT DOMAIN 101 104 POLY-ALA.
FT DNA_BIND 126 217 FORK-HEAD.
FT DOMAIN 247 250 POLY-ALA.
FT DOMAIN 296 306 POLY-ALA.
FT DOMAIN 398 409 POLY-GLY.
FT DOMAIN 421 426 POLY-GLY.
FT DOMAIN 442 445 POLY-ALA.
SQ SEQUENCE 497 AA; 49007 MW; EAAFA98D216BE019 CRC64;
Query Match 30.6%; Score 59; DB 1; Length 497;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
QY 4 PT--LRQWLAARAGGGKGGG 22
| | | | | | | | | | | | | |
DB 385 PTALLRQGLKTDAGGGAGGG 405
| | | | | | | | | | | | | |

RESULT 5
INS_BOVIN STANDARD; PRT; 105 AA.
AC P01317;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
OS INS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88288209; PubMed=2456452;
RA D'Agostino J., Younes M.A., White J.W., Besch P.K., Field J.B.,
RA Frazier M.L.;
RT "Cloning and nucleotide sequence analysis of complementary
RT deoxyribonucleic acid for bovine preproinsulin.";
RL Mol. Endocrinol. 1:327-331(1987).
RN [2]
RP SEQUENCE OF 25-105.
RX MEDLINE=71166442; PubMed=4928892;
RA Nolan C., Margoliash E., Peterson J.D., Steiner D.F.;
RT "The structure of bovine proinsulin.";
RL J. Biol. Chem. 246:2780-2795(1971).
RN [3]
RP SEQUENCE OF 25-54.
RA Sanger F., Tuppy H.;
RT "The amino-acid sequence in the phenylalanyl chain of insulin. 2. The
RT investigation of peptides from enzymic hydrolysates.";
RL Biochem. J. 49:481-490(1951).
RN [4]
RP SEQUENCE OF 57-82.
RX MEDLINE=7116409; PubMed=5545080;
RA Steiner D.F., Cho S., Oyer P.E., Terris S., Peterson J.D.,
RA Rubenstein A.H.;
RT "Isolation and characterization of proinsulin C-peptide from bovine
RT pancreas.";
RL J. Biol. Chem. 246:1365-1374(1971).
RN [5]
RP SEQUENCE OF 57-82.

```

RX MEDLINE=71257721; PubMed=5105368;
RA Salokangas A., Smyth D.G., Markussen J., Sundby F.;
RT "Bovine proinsulin: amino acid sequence of the C-peptide isolated
from pancreas.";
RN Eur. J. Biochem. 20:183-189(1971).
RN [6]
RP SEQUENCE OF 85-105.
RA Sanger F., Thompson E.O.P.;
RT "The amino-acid sequence in the glycol chain of insulin. 2. The
RT investigation of peptides from enzymic hydrolysates.";
RL Biochem. J. 53:366-374(1953).
RN [7]
RN AMIDES, SEQUENCE OF 25-54 AND 85-105, AND DISULFIDE BONDS.
RA Kyle A.P., Sanger F., Smith L.F., Kitai R.;
RT "The disulphide bonds of insulin.";
RL Biochem. J. 60:541-556(1955).
RN [8]
RN X-RAY CRYSTALLOGRAPHY.
RA Smith G.D., Duax W.L., Dodson E.J., Dodson G.G., de Graaf R.A.G.,
RA Reynolds C.D.;
RT "The structure of des-Phe b1 bovine insulin.";
RL Acta Crystallogr. B 38:3028-3032(1982).
RN [9]
RN X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).
RX MEDLINE=97285914; PubMed=9141131;
RA Brange J., Dodson G.G., Edwards D.J., Holden P.H., Whittingham J.L.;
RT "A model of insulin fibrils derived from the X-ray crystal structure
of a monomeric insulin (despentapeptide insulin).";
RL Proteins 27:507-516(1997).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- DATABASE: NAME-Protein Spotlight;
CC NOTE=Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sptlt009.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M54979; AAA30722.1; -;
CC PIR; A01585; IPBO.
CC PIR; A40909; A40909.
CC PDB; 2INS; 31-MAY-84.
CC PDB; IAPH; 31-OCT-93.
CC PDB; IBPH; 31-OCT-93.
CC PDB; ICPH; 31-OCT-93.
CC PDB; IDPH; 31-OCT-93.
CC PDB; IPID; 07-DEC-96.
CC InterPro; IPR000739; Insulin_IGF_relaxin.
CC Pfam; PF00049; Insulin; 1.
CC PRINTS; PR00276; INSULIN.
CC PRINTS; PR00277; INSULIN.
CC SMART; SM00078; IIGF; 1.
CC PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 82 C PEPTIDE.
FT CHAIN 85 105 INSULIN A CHAIN.
FT DISULFID 31 91 INTERCHAIN.
FT DISULFID 43 104 INTERCHAIN.
FT DISULFID 90 95
FT TURN 32 32

FT HELIX 33 46
FT STRAND 48 48
FT HELIX 86 90
FT TURN 91 94
FT HELIX 97 101
FT TURN 102 103
FT STRAND 104 104
SQ SEQUENCE 105 AA; 11393 MW; 75307CF78E61C06A CRC64;

Query Match 30.3%; Score 58.5; DB 1; Length 105;
Best Local Similarity 50.0%; Pred. No. 4.7;
Matches 13; Conservative 2; Mismatches 8; Indels 3; Gaps 1;

QY 1 IEQPTLRQWLAAARAGGKGKGIEGP 26
DB 58 VEGP--QVGALELAGGPGAGGLEGP 80

RESULT 6
INS_SHEEP
ID INS_SHEEP STANDARD; PRT; 105 AA.
AC P01318;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin precursor.
GN INS.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94280618; PubMed=8011164;
RA Ohlssen S.M., Lugenbeel K.A., Wong E.A.;
RT "Characterization of the linked ovine insulin and insulin-like growth
RT factor-II genes.";
RL DNA Cell Biol. 13:377-388(1994).
RN [2]
RP SEQUENCE OF 25-54 AND 85-105.
RA Brown H., Sanger F., Kitai R.;
RT "The structure of pig and sheep insulins.";
RL Biochem. J. 60:556-565(1955).
RN [3]
RP SEQUENCE OF 57-82.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00659; AAB60625.1; -;
CC PIR; S16430; INSH.
CC HSSP; P01315; 9INS.
CC InterPro; IPR000739; Insulin_IGF_relaxin.
CC DR

DR	TRANSFAC; T01755; -	
DR	MGI; 96210; Hoxd9.	
DR	InterPro; IPR001356; Homeobox.	
DR	Pfam; PF00046; homeobox; 1.	
DR	PRINTS; PR00024; HOMEBOX.	
DR	SMART; SM00389; HOX; 1.	
DR	PROSITE; PS00027; HOMEBOX_1; 1.	
DR	PROSITE; PS50071; HOMEBOX_2; 1.	
KW	Homeobox; DNA-binding; Developmental protein; Nuclear protein;	
KW	Transcription regulation.	
FT	DOMAIN 113 140	GLY-RICH.
FT	DOMAIN 119 131	POLY-GLY.
FT	DOMAIN 163 176	SER/THR-RICH.
FT	DOMAIN 272 331	HOMEBOX.
FT	DNA_BIND 272 331	HOMEBOX.
DR	SEQUENCE 339 AA; 34992 MW; 370DC47C6929F7E1 CRC64;	
QY	3 GPTLRQWL-----AARAGGKGGGTGGP 26	
DB	101 GRVYSWMEPLPGFPGAGGGGGGGGGGGP 132	
RESULT 8		
SC02_HUMAN	STANDARD; PRT; 266 AA.	
ID	SC02_HUMAN	
AC	O43819; Q9UK97;	
DT	30-MAY-2000 (Rel. 39, Created)	
DD	30-MAY-2000 (Rel. 39, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	SC02 protein homolog, mitochondrial precursor.	
DE	SC02.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OC	NCBI_TaxID=9606;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RP	TISSUE=Monocytes;	
RC	Smink L.J., Burton J.;	
RC	Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.	
RRL	[2]	
RP	SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.	
RP	MEDLINE=20014747; PubMed=10545952;	
RX	Papadopolou L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,	
RA	Sadlock J.E., Krishna S., Walker W., Selby J., Glerum D.M.,	
RA	Van Coster R., Lyon G., Scallan E., Lebel R., Kaplan P., Shanske S.,	
RA	De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;	
RA	"Fetal infantile cardioencephalomyopathy with COX deficiency and	
RA	mutations in SC02, a COX assembly gene.";	
RA	Nat. Genet. 23:333-337(1999).	
CC	!- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER	
CC	!- TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.	
CC	!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).	
CC	!- TISSUE SPECIFICITY: UBIQUITOUS.	
CC	!- DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE	
CC	CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS	
CC	CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND	
CC	GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX	
CC	ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX	
CC	DEFICIENCIES.	
CC	!- SIMILARITY: BELONGS TO THE SC01/2 FAMILY.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	

DR EMBL; AF177385; AAF05313.1; -;
DR EMBL; AL021683; CAA16671.1; -;
DR MIM; 604272; -;
DR MIM; 604377; -;
DR MIM; 220110; -;
DR InterPro; IPR003782; SC01_Senc.
DR Pfam; PF02630; SC01_Senc; 1.
KW Mitochondrion; Transist peptide; Disease mutation; Polymorphism.
FT TRANSIT 1 41 MITOCHONDRION (POTENTIAL).
FT CHAIN 42 266 SC02 PROTEIN HOMOLOG.
FT VARIANT 20 20 R -> P (IN DBSNP:140523).
FT VARIANT 140 140 /FTIQ-VAR_0111738.
FT VARIANT 225 225 E -> K (IN PIC).
FT VARIANT 225 225 S -> F (IN PIC).
FT SEQUENCE 266 AA; 29810 MW; BC2F40E057329BF3 CRC64;
Query Match 29.8%; Score 57.5; DB 1; Length 266;
Best Local Similarity 35.4%; Pred. No. 14;
Matches 17; Conservative 3; Mismatches 11; Indels 17; Gaps 2;
QY 6 LROWLAARAGGK-GGGGEGFTLR-----QWLAARA 36
DB 33 LRSWLLSROGPAETGGGQGPGLRTRLLTGLFGAGLGAWLARA 80
RESULT 9
FZDB_HUMAN STANDARD; PRT; 694 AA.
AC Q9H461;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Frizzled 8 precursor (Frizzled-8) (Fz-8) (hFz8).
GN FZD8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=21192958; PubMed=11295046;
RA Saitoh T., Hirai M., Katoh M.;
RT "Molecular cloning and characterization of human Frizzled-8 gene on
chromosome 10p11.2";
RL Int. J. Oncol. 18:991-996(2001).
[2]
SEQUENCE FROM N.A.
RA Heath P.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors
are coupled to the beta-catenin canonical signaling pathway, which
leads to the activation of dishevelled proteins, inhibition of GSK-
3 kinase, nuclear accumulation of beta-catenin and activation of
Wnt target genes. A second signaling pathway involving PKC and
calcium fluxes has been seen for some family members, but it is
not yet clear if it represents a distinct pathway or if it can be
integrated in the canonical pathway, as PKC seems to be required
for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem
to involve interactions with G-proteins. May be involved in
transduction and intercellular transmission of polarity
information during tissue morphogenesis and/or in differentiated
tissues.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Most abundant in fetal kidney, followed by
brain and lung. In adult tissues, expressed in kidney, heart,
pancreas and skeletal muscle.
CC -1- DOMAIN: Lys-Thr-X-X-X-Trp motif is involved in the activation of
the Wnt/beta-catenin signaling pathway (By similarity).
CC -1- DOMAIN: The fz domain is involved in binding with Wnt ligands (by
similarity).
CC -1- SIMILARITY: BELONGS TO FAMILY FZ/SMO OF G-PROTEIN COUPLED

RECEPTORS.
-1- SIMILARITY: CONTAINS 1 FRIZZLED (FZ) DOMAIN.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; AB043703; BAB41064.1; -;
EMBL; AL21749; CAC10185.1; -;
MIM; 606146; -;
InterPro; IPR000539; Frizzled.
InterPro; IPR000024; Fz_domain.
InterPro; IPR000832; GPCR_secretin.
Pfam; PF01534; Frizzled; 1.
Pfam; PF01392; Fz; 1.
PRINTS; PR00489; FRIZZLED.
PROSITE; PS00063; FRI; 1.
PROSITE; PS50038; Fz; 1.
Multigene family; G-PROTEIN_RECEP_F2_4; 1.
Developmental protein; Glycoprotein; Signal.
SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 694 FRIZZLED 8.
FT DOMAIN 28 275 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 276 296 1 (POTENTIAL).
FT DOMAIN 297 312 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 313 333 2 (POTENTIAL).
FT DOMAIN 334 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 417 3 (POTENTIAL).
FT DOMAIN 418 439 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 440 460 4 (POTENTIAL).
FT DOMAIN 461 483 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 484 504 5 (POTENTIAL).
FT DOMAIN 505 532 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 533 553 6 (POTENTIAL).
FT DOMAIN 554 584 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 585 605 7 (POTENTIAL).
FT DOMAIN 606 694 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 30 151 FZ.
FT DOMAIN 168 172 POLY-PRO.
FT DOMAIN 194 202 POLY-GLY.
FT DOMAIN 211 216 POLY-GLY.
FT DOMAIN 639 663 POLY-GLY.
FT SITE 608 613 LYS-THR-X-X-X-TRP MOTIF.
FT SITE 692 694 PDZ-BINDING.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 694 AA; 73300 MW; E740CBFDA2A233EF CRC64;
Query Match 29.5%; Score 57; DB 1; Length 694;
Best Local Similarity 32.6%; Pred. No. 37;
Matches 15; Conservative 0; Mismatches 9; Indels 22; Gaps 1;
QY 3 GPTLRQW-----LAARAGGGGGGIEGP 26
DB 607 GKTLESWRSILTRCCWASKGAAGVGGAGATAAGGGGPGGGGGP 652
RESULT 10
GAG_MMTVB STANDARD; PRT; 591 AA.
ID GAG_MMTVB
AC P10258;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE GAG polypeptide [Contains: Protein P10; Phosphorylated protein pp21;
Protein P3; Protein P8; Major core protein P27; Nucleic acid binding
protein P14].

DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA polymerase I (EC 2.7.7.7) (POL I).
 GN POLA OR RV1629 OR WT1665 OR WTCY01B2.21.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1773;
 RX SEQUENCE FROM N.A.
 RP STRAIN=H37RV;
 RC MEDLINE=94124016; PubMed=8294019;
 RA Mizrahi V., Huberts P., Dawes S.S., Dudding L.R.;
 RT "A PCR method for the sequence analysis of the *gyrA*, *polA* and *rnha*
 RL gene segments from mycobacteria.";
 RN Gene 136:287-290(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RL complete genome sequence.";
 RN Nature 393:537-544(1998).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains.";
 RN Submitted (APR-2001); to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
 CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + (DNA)(N).
 CC -1- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L11920; AAB46393.1; -;
 DR EMBL; Z95554; CAB08882.1; -;
 DR EMBL; AF007030; AAK45935.1; -;
 DR HSSP; P19821; 1BGX.
 DR TIGR; MT1665; -;
 DR TubercuList; RV1629; -;
 DR InterPro; IPR002562; 3_5_exonuclease.
 DR InterPro; IPR002421; 5_3_exonuclease.
 DR InterPro; IPR002298; DNA_poli.
 DR InterPro; IPR001098; DNA_poli_A.
 DR InterPro; IPR000513; EXO_N1.
 DR InterPro; IPR003583; HHH1.
 DR InterPro; IPR003584; HHH2.
 DR Pfam; PF01367; 5_3_exonuclease; 1.
 DR Pfam; PF02739; 5_3_exonuc_N; 1.

DR Pfam; PF00476; DNA_poli_A; 1.
 DR PRINTS; PR00868; DNAPOLI.
 DR SMART; SM00474; 35EXOC; 1.
 DR SMART; SM00475; 53EXOC; 1.
 DR SMART; SM00278; HHH1; 1.
 DR SMART; SM00279; HHH2; 1.
 DR SMART; SM00482; POLAC; 1.
 DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
 KW Hydrolase; Exonuclease; DNA-binding; Complete proteome.
 SQ SEQUENCE 904 AA; 98471 MW; 1C8E560FE5F74323 CRC64;
 Query Match 28.5%; Score 55; DB 1; Length 904;
 Best Local Similarity 58.8%; Pred. No. 75;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 5 TLQWLARAGGKGGG 21
 :||||| :| :|
 Db 325 TVRQWLAEHAGDGRAG 341
 RESULT 15
 LSR2_MYCLE
 ID LSR2_MYCLE STANDARD; PRT; 112 AA.
 AC P24094;
 DT 01-NAR-1992 (Rel. 21, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE LSR2 protein precursor (15 kDa antigen) (A15).
 GN LSR2 OR ML0234.
 OS Mycobacterium leprae.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1769;
 RX MEDLINE=92040086; PubMed=1840579;
 RA Sela S., Thole J.E., Ottenhoff T.H., Clark-Curtiss J.E.;
 RT "Identification of Mycobacterium leprae antigens from a cosmid
 RT library: characterization of a 15-kilodalton antigen that is
 RT recognized by both the humoral and cellular immune systems in leprosy
 RT patients.";
 RN Infect. Immun. 59:4117-4124(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 [3]
 RP SEQUENCE OF 24-115 FROM N.A.
 RX MEDLINE=91126054; PubMed=1992456;
 RA Laal S., Sharma Y.D., Prasad H.K., Murtaza A., Singh S., Tangri S.,
 RA Misra R.S., Nath I.;
 RT "Recombinant fusion protein identified by lepromatous sera mimics
 RT native Mycobacterium leprae in T-cell responses across the leprosy
 RT spectrum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:1054-1058(1991).
 CC -1- FUNCTION: DOMINANT T-CELL ANTIGEN AND STIMULATES
 CC LYMPHOPROLIFERATION.
 CC -1- DISEASE: MOST PROBABLY CAUSES THE LYMPHOPROLIFERATIVE RESPONSES
 CC OCCURRING IN LEPROSY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

Db 145 GPTRNPWVDRVPGSGG 163
|||||
RESULT 18
NUMB_DROME
ID NUMB_DROME STANDARD; PRT; 556 AA.
AC P16554;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Numb protein.
GN NUMB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89324081; PubMed=2752427;
RA Uemura T., Shepherd S., Ackerman L., Jan L.Y., Jan Y.N.;
RT "Numb, a gene required in determination of cell fate during sensory
organ formation in Drosophila embryos.";
RL cell 58:349-360(1989).
RN [2]
RP STRUCTURE BY NMR OF 64-210.
RX MEDLINE=9061335; PubMed=9846878;
RA Li S.-C., Zwaehlen C., Vincent S.J., McGlade C.J., Kay L.E., Pawson T.,
RA Forman-Kay J.D.;
RT "Structure of a Numb PTB domain-peptide complex suggests a basis for
diverse binding specificity.";
RL Nat. Struct. Biol. 5:1075-1083(1998).
CC -1- FUNCTION: NUMB IS REQUIRED IN DETERMINATION OF CELL FATE DURING
SENSORY ORGAN FORMATION IN DROSOPHILA EMBRYOS. IT FUNCTIONS IN
NUCLEI AND SEEMS TO INTERACT WITH NUCLEIC ACIDS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 PID DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M27815; AAA28730.1; -.
DR PIR; A32466; A32466.
DR PDB; 2NMB; 04-NOV-98.
DR FlyBase; FBgn0002973; numb.
DR InterPro; IPR000050; PID_domain.
DR Pfam; PF00640; PID; 1.
DR SMART; SM00462; PTB; 1.
DR PROSITE; PS01179; PID; 1.
DR Nucleic acid binding; Alternative initiation; 3D-structure.
FT CHAIN 1 556
FT CHAIN 42 556
FT NUMB PROTEIN, MATERNAL ISOFORM
FT (PROBABLE).
FT INIT_MET 42 42
FT NP_BIND 22 29
FT DOMAIN 25 57
FT DOMAIN 81 208
FT PID.
SQ SEQUENCE 556 AA; 60628 MW; 4FECAAE9C98FEE71 CRC64;

Query Match 28.0%; Score 54; DB 1; Length 556;
Best Local Similarity 42.3%; Pred. No. 62;
Matches 11; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 8 OWLAARAGGKGGGEGPTLRQWLA 33
|||||
Db 486 QTLASGTGAAGGGGPDPPFAEWYA 511
|||||

```

```

RESULT 19
TBX2_HUMAN
ID TBX2_HUMAN STANDARD; PRT; 702 AA.
AC Q13207; Q16424;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-box transcription factor TBX2 (T-box protein 2).
GN TBX2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Fetal kidney;
RX MEDLINE=96015055; PubMed=8530034;
RA Campbell C., Goodrich K., Casey G., Beatty B.;
RT "Cloning and mapping of a human gene (TBX2) sharing a highly conserved
protein motif with the Drosophila omb gene.";
RL Genomics 28:255-260(1995).
RN [2]
RP SEQUENCE OF 152-245 FROM N.A.
RX TISSUE=Fetal kidney;
RX MEDLINE=96169568; PubMed=8597636;
RA Law D.J., Gebuhr T., Garvey N., Agulnik S.I., Silver L.M.;
RT "Identification, characterization, and localization to chromosome
17q21-22 of the human TBX2 homolog, member of a conserved
developmental gene family.";
RL Mamm. Genome 6:793-797(1995).
CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES
REQUIRED FOR MESODERM DIFFERENTIATION. PROBABLY PLAYS A ROLE IN
LIMB PATTERN FORMATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ADULT IN KIDNEY, LUNG,
AND PLACENTA. WEAK EXPRESSION IN HEART AND OVARY.
CC -1- SIMILARITY: CONTAINS 1 T-BOX DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U28049; AAA73861.1; -.
DR EMBL; S81264; AAB36216.1; -.
DR HSP; P24781; LXBR.
DR MIM; 600747; -.
DR InterPro; IPR001699; T-box.
DR Pfam; PF00907; T-box; 1.
DR PRINTS; PR00937; TBOX.
DR SMART; SM00425; TBOX; 1.
DR PROSITE; PS01283; TBOX_1; 1.
DR PROSITE; PS01264; TBOX_2; 1.
DR PROSITE; PS01264; TBOX_3; 1.
DR PROSITE; PS01264; TBOX_4; 1.
DR PROSITE; PS01264; TBOX_5; 1.
DR PROSITE; PS01264; TBOX_6; 1.
DR PROSITE; PS01264; TBOX_7; 1.
DR PROSITE; PS01264; TBOX_8; 1.
DR PROSITE; PS01264; TBOX_9; 1.
DR PROSITE; PS01264; TBOX_10; 1.
DR PROSITE; PS01264; TBOX_11; 1.
DR PROSITE; PS01264; TBOX_12; 1.
DR PROSITE; PS01264; TBOX_13; 1.
DR PROSITE; PS01264; TBOX_14; 1.
DR PROSITE; PS01264; TBOX_15; 1.
DR PROSITE; PS01264; TBOX_16; 1.
DR PROSITE; PS01264; TBOX_17; 1.
DR PROSITE; PS01264; TBOX_18; 1.
DR PROSITE; PS01264; TBOX_19; 1.
DR PROSITE; PS01264; TBOX_20; 1.
DR PROSITE; PS01264; TBOX_21; 1.
DR PROSITE; PS01264; TBOX_22; 1.
DR PROSITE; PS01264; TBOX_23; 1.
DR PROSITE; PS01264; TBOX_24; 1.
DR PROSITE; PS01264; TBOX_25; 1.
DR PROSITE; PS01264; TBOX_26; 1.
DR PROSITE; PS01264; TBOX_27; 1.
DR PROSITE; PS01264; TBOX_28; 1.
DR PROSITE; PS01264; TBOX_29; 1.
DR PROSITE; PS01264; TBOX_30; 1.
DR PROSITE; PS01264; TBOX_31; 1.
DR PROSITE; PS01264; TBOX_32; 1.
DR PROSITE; PS01264; TBOX_33; 1.
DR PROSITE; PS01264; TBOX_34; 1.
DR PROSITE; PS01264; TBOX_35; 1.
DR PROSITE; PS01264; TBOX_36; 1.
DR PROSITE; PS01264; TBOX_37; 1.
DR PROSITE; PS01264; TBOX_38; 1.
DR PROSITE; PS01264; TBOX_39; 1.
DR PROSITE; PS01264; TBOX_40; 1.
DR PROSITE; PS01264; TBOX_41; 1.
DR PROSITE; PS01264; TBOX_42; 1.
DR PROSITE; PS01264; TBOX_43; 1.
DR PROSITE; PS01264; TBOX_44; 1.
DR PROSITE; PS01264; TBOX_45; 1.
DR PROSITE; PS01264; TBOX_46; 1.
DR PROSITE; PS01264; TBOX_47; 1.
DR PROSITE; PS01264; TBOX_48; 1.
DR PROSITE; PS01264; TBOX_49; 1.
DR PROSITE; PS01264; TBOX_50; 1.
DR PROSITE; PS01264; TBOX_51; 1.
DR PROSITE; PS01264; TBOX_52; 1.
DR PROSITE; PS01264; TBOX_53; 1.
DR PROSITE; PS01264; TBOX_54; 1.
DR PROSITE; PS01264; TBOX_55; 1.
DR PROSITE; PS01264; TBOX_56; 1.
DR PROSITE; PS01264; TBOX_57; 1.
DR PROSITE; PS01264; TBOX_58; 1.
DR PROSITE; PS01264; TBOX_59; 1.
DR PROSITE; PS01264; TBOX_60; 1.
DR PROSITE; PS01264; TBOX_61; 1.
DR PROSITE; PS01264; TBOX_62; 1.
DR PROSITE; PS01264; TBOX_63; 1.
DR PROSITE; PS01264; TBOX_64; 1.
DR PROSITE; PS01264; TBOX_65; 1.
DR PROSITE; PS01264; TBOX_66; 1.
DR PROSITE; PS01264; TBOX_67; 1.
DR PROSITE; PS01264; TBOX_68; 1.
DR PROSITE; PS01264; TBOX_69; 1.
DR PROSITE; PS01264; TBOX_70; 1.
DR PROSITE; PS01264; TBOX_71; 1.
DR PROSITE; PS01264; TBOX_72; 1.
DR PROSITE; PS01264; TBOX_73; 1.
DR PROSITE; PS01264; TBOX_74; 1.
DR PROSITE; PS01264; TBOX_75; 1.
DR PROSITE; PS01264; TBOX_76; 1.
DR PROSITE; PS01264; TBOX_77; 1.
DR PROSITE; PS01264; TBOX_78; 1.
DR PROSITE; PS01264; TBOX_79; 1.
DR PROSITE; PS01264; TBOX_80; 1.
DR PROSITE; PS01264; TBOX_81; 1.
DR PROSITE; PS01264; TBOX_82; 1.
DR PROSITE; PS01264; TBOX_83; 1.
DR PROSITE; PS01264; TBOX_84; 1.
DR PROSITE; PS01264; TBOX_85; 1.
DR PROSITE; PS01264; TBOX_86; 1.
DR PROSITE; PS01264; TBOX_87; 1.
DR PROSITE; PS01264; TBOX_88; 1.
DR PROSITE; PS01264; TBOX_89; 1.
DR PROSITE; PS01264; TBOX_90; 1.
DR PROSITE; PS01264; TBOX_91; 1.
DR PROSITE; PS01264; TBOX_92; 1.
DR PROSITE; PS01264; TBOX_93; 1.
DR PROSITE; PS01264; TBOX_94; 1.
DR PROSITE; PS01264; TBOX_95; 1.
DR PROSITE; PS01264; TBOX_96; 1.
DR PROSITE; PS01264; TBOX_97; 1.
DR PROSITE; PS01264; TBOX_98; 1.
DR PROSITE; PS01264; TBOX_99; 1.
DR PROSITE; PS01264; TBOX_100; 1.
DR PROSITE; PS01264; TBOX_101; 1.
DR PROSITE; PS01264; TBOX_102; 1.
DR PROSITE; PS01264; TBOX_103; 1.
DR PROSITE; PS01264; TBOX_104; 1.
DR PROSITE; PS01264; TBOX_105; 1.
DR PROSITE; PS01264; TBOX_106; 1.
DR PROSITE; PS01264; TBOX_107; 1.
DR PROSITE; PS01264; TBOX_108; 1.
DR PROSITE; PS01264; TBOX_109; 1.
DR PROSITE; PS01264; TBOX_110; 1.
DR PROSITE; PS01264; TBOX_111; 1.
DR PROSITE; PS01264; TBOX_112; 1.
DR PROSITE; PS01264; TBOX_113; 1.
DR PROSITE; PS01264; TBOX_114; 1.
DR PROSITE; PS01264; TBOX_115; 1.
DR PROSITE; PS01264; TBOX_116; 1.
DR PROSITE; PS01264; TBOX_117; 1.
DR PROSITE; PS01264; TBOX_118; 1.
DR PROSITE; PS01264; TBOX_119; 1.
DR PROSITE; PS01264; TBOX_120; 1.
DR PROSITE; PS01264; TBOX_121; 1.
DR PROSITE; PS01264; TBOX_122; 1.
DR PROSITE; PS01264; TBOX_123; 1.
DR PROSITE; PS01264; TBOX_124; 1.
DR PROSITE; PS01264; TBOX_125; 1.
DR PROSITE; PS01264; TBOX_126; 1.
DR PROSITE; PS01264; TBOX_127; 1.
DR PROSITE; PS01264; TBOX_128; 1.
DR PROSITE; PS01264; TBOX_129; 1.
DR PROSITE; PS01264; TBOX_130; 1.
DR PROSITE; PS01264; TBOX_131; 1.
DR PROSITE; PS01264; TBOX_132; 1.
DR PROSITE; PS01264; TBOX_133; 1.
DR PROSITE; PS01264; TBOX_134; 1.
DR PROSITE; PS01264; TBOX_135; 1.
DR PROSITE; PS01264; TBOX_136; 1.
DR PROSITE; PS01264; TBOX_137; 1.
DR PROSITE; PS01264; TBOX_138; 1.
DR PROSITE; PS01264; TBOX_139; 1.
DR PROSITE; PS01264; TBOX_140; 1.
DR PROSITE; PS01264; TBOX_141; 1.
DR PROSITE; PS01264; TBOX_142; 1.
DR PROSITE; PS01264; TBOX_143; 1.
DR PROSITE; PS01264; TBOX_144; 1.
DR PROSITE; PS01264; TBOX_145; 1.
DR PROSITE; PS01264; TBOX_146; 1.
DR PROSITE; PS01264; TBOX_147; 1.
DR PROSITE; PS01264; TBOX_148; 1.
DR PROSITE; PS01264; TBOX_149; 1.
DR PROSITE; PS01264; TBOX_150; 1.
DR PROSITE; PS01264; TBOX_151; 1.
DR PROSITE; PS01264; TBOX_152; 1.
DR PROSITE; PS01264; TBOX_153; 1.
DR PROSITE; PS01264; TBOX_154; 1.
DR PROSITE; PS01264; TBOX_155; 1.
DR PROSITE; PS01264; TBOX_156; 1.
DR PROSITE; PS01264; TBOX_157; 1.
DR PROSITE; PS01264; TBOX_158; 1.
DR PROSITE; PS01264; TBOX_159; 1.
DR PROSITE; PS01264; TBOX_160; 1.
DR PROSITE; PS01264; TBOX_161; 1.
DR PROSITE; PS01264; TBOX_162; 1.
DR PROSITE; PS01264; TBOX_163; 1.
DR PROSITE; PS01264; TBOX_164; 1.
DR PROSITE; PS01264; TBOX_165; 1.
DR PROSITE; PS01264; TBOX_166; 1.
DR PROSITE; PS01264; TBOX_167; 1.
DR PROSITE; PS01264; TBOX_168; 1.
DR PROSITE; PS01264; TBOX_169; 1.
DR PROSITE; PS01264; TBOX_170; 1.
DR PROSITE; PS01264; TBOX_171; 1.
DR PROSITE; PS01264; TBOX_172; 1.
DR PROSITE; PS01264; TBOX_173; 1.
DR PROSITE; PS01264; TBOX_174; 1.
DR PROSITE; PS01264; TBOX_175; 1.
DR PROSITE; PS01264; TBOX_176; 1.
DR PROSITE; PS01264; TBOX_177; 1.
DR PROSITE; PS01264; TBOX_178; 1.
DR PROSITE; PS01264; TBOX_179; 1.
DR PROSITE; PS01264; TBOX_180; 1.
DR PROSITE; PS01264; TBOX_181; 1.
DR PROSITE; PS01264; TBOX_182; 1.
DR PROSITE; PS01264; TBOX_183; 1.
DR PROSITE; PS01264; TBOX_184; 1.
DR PROSITE; PS01264; TBOX_185; 1.
DR PROSITE; PS01264; TBOX_186; 1.
DR PROSITE; PS01264; TBOX_187; 1.
DR PROSITE; PS01264; TBOX_188; 1.
DR PROSITE; PS01264; TBOX_189; 1.
DR PROSITE; PS01264; TBOX_190; 1.
DR PROSITE; PS01264; TBOX_191; 1.
DR PROSITE; PS01264; TBOX_192; 1.
DR PROSITE; PS01264; TBOX_193; 1.
DR PROSITE; PS01264; TBOX_194; 1.
DR PROSITE; PS01264; TBOX_195; 1.
DR PROSITE; PS01264; TBOX_196; 1.
DR PROSITE; PS01264; TBOX_197; 1.
DR PROSITE; PS01264; TBOX_198; 1.
DR PROSITE; PS01264; TBOX_199; 1.
DR PROSITE; PS01264; TBOX_200; 1.
DR PROSITE; PS01264; TBOX_201; 1.
DR PROSITE; PS01264; TBOX_202; 1.
DR PROSITE; PS01264; TBOX_203; 1.
DR PROSITE; PS01264; TBOX_204; 1.
DR PROSITE; PS01264; TBOX_205; 1.
DR PROSITE; PS01264; TBOX_206; 1.
DR PROSITE; PS01264; TBOX_207; 1.
DR PROSITE; PS01264; TBOX_208; 1.
DR PROSITE; PS01264; TBOX_209; 1.
DR PROSITE; PS01264; TBOX_210; 1.
DR PROSITE; PS01264; TBOX_211; 1.
DR PROSITE; PS01264; TBOX_212; 1.
DR PROSITE; PS01264; TBOX_213; 1.
DR PROSITE; PS01264; TBOX_214; 1.
DR PROSITE; PS01264; TBOX_215; 1.
DR PROSITE; PS01264; TBOX_216; 1.
DR PROSITE; PS01264; TBOX_217; 1.
DR PROSITE; PS01264; TBOX_218; 1.
DR PROSITE; PS01264; TBOX_219; 1.
DR PROSITE; PS01264; TBOX_220; 1.
DR PROSITE; PS01264; TBOX_221; 1.
DR PROSITE; PS01264; TBOX_222; 1.
DR PROSITE; PS01264; TBOX_223; 1.
DR PROSITE; PS01264; TBOX_224; 1.
DR PROSITE; PS01264; TBOX_225; 1.
DR PROSITE; PS01264; TBOX_226; 1.
DR PROSITE; PS01264; TBOX_227; 1.
DR PROSITE; PS01264; TBOX_228; 1.
DR PROSITE; PS01264; TBOX_229; 1.
DR PROSITE; PS01264; TBOX_230; 1.
DR PROSITE; PS01264; TBOX_231; 1.
DR PROSITE; PS01264; TBOX_232; 1.
DR PROSITE; PS01264; TBOX_233; 1.
DR PROSITE; PS01264; TBOX_234; 1.
DR PROSITE; PS01264; TBOX_235; 1.
DR PROSITE; PS01264; TBOX_236; 1.
DR PROSITE; PS01264; TBOX_237; 1.
DR PROSITE; PS01264; TBOX_238; 1.
DR PROSITE; PS01264; TBOX_239; 1.
DR PROSITE; PS01264; TBOX_240; 1.
DR PROSITE; PS01264; TBOX_241; 1.
DR PROSITE; PS01264; TBOX_242; 1.
DR PROSITE; PS01264; TBOX_243; 1.
DR PROSITE; PS01264; TBOX_244; 1.
DR PROSITE; PS01264; TBOX_245; 1.
DR PROSITE; PS01264; TBOX_246; 1.
DR PROSITE; PS01264; TBOX_247; 1.
DR PROSITE; PS01264; TBOX_248; 1.
DR PROSITE; PS01264; TBOX_249; 1.
DR PROSITE; PS01264; TBOX_250; 1.
DR PROSITE; PS01264; TBOX_251; 1.
DR PROSITE; PS01264; TBOX_252; 1.
DR PROSITE; PS01264; TBOX_253; 1.
DR PROSITE; PS01264; TBOX_254; 1.
DR PROSITE; PS01264; TBOX_255; 1.
DR PROSITE; PS01264; TBOX_256; 1.
DR PROSITE; PS01264; TBOX_257; 1.
DR PROSITE; PS01264; TBOX_258; 1.
DR PROSITE; PS01264; TBOX_259; 1.
DR PROSITE; PS01264; TBOX_260; 1.
DR PROSITE; PS01264; TBOX_261; 1.
DR PROSITE; PS01264; TBOX_262; 1.
DR PROSITE; PS01264; TBOX_263; 1.
DR PROSITE; PS01264; TBOX_264; 1.
DR PROSITE; PS01264; TBOX_265; 1.
DR PROSITE; PS01264; TBOX_266; 1.
DR PROSITE; PS01264; TBOX_267; 1.
DR PROSITE; PS01264; TBOX_268; 1.
DR PROSITE; PS01264; TBOX_269; 1.
DR PROSITE; PS01264; TBOX_270; 1.
DR PROSITE; PS01264; TBOX_271; 1.
DR PROSITE; PS01264; TBOX_272; 1.
DR PROSITE; PS01264; TBOX_273; 1.
DR PROSITE; PS01264; TBOX_274; 1.
DR PROSITE; PS01264; TBOX_275; 1.
DR PROSITE; PS01264; TBOX_276; 1.
DR PROSITE; PS01264; TBOX_277; 1.
DR PROSITE; PS01264; TBOX_278; 1.
DR PROSITE; PS01264; TBOX_279; 1.
DR PROSITE; PS01264; TBOX_280; 1.
DR PROSITE; PS01264; TBOX_281; 1.
DR PROSITE; PS01264; TBOX_282; 1.
DR PROSITE; PS01264; TBOX_283; 1.
DR PROSITE; PS01264; TBOX_284; 1.
DR PROSITE; PS01264; TBOX_285; 1.
DR PROSITE; PS01264; TBOX_286; 1.
DR PROSITE; PS01264; TBOX_287; 1.
DR PROSITE; PS01264; TBOX_288; 1.
DR PROSITE; PS01264; TBOX_289; 1.
DR PROSITE; PS01264; TBOX_290; 1.
DR PROSITE; PS01264; TBOX_291; 1.
DR PROSITE; PS01264; TBOX_292; 1.
DR PROSITE; PS01264; TBOX_293; 1.
DR PROSITE; PS01264; TBOX_294; 1.
DR PROSITE; PS01264; TBOX_295; 1.
DR PROSITE; PS01264; TBOX_296; 1.
DR PROSITE; PS01264; TBOX_297; 1.
DR PROSITE; PS01264; TBOX_298; 1.
DR PROSITE; PS01264; TBOX_299; 1.
DR PROSITE; PS01264; TBOX_300; 1.
DR PROSITE; PS01264; TBOX_301; 1.
DR PROSITE; PS01264; TBOX_302; 1.
DR PROSITE; PS01264; TBOX_303; 1.
DR PROSITE; PS01264; TBOX_304; 1.
DR PROSITE; PS01264; TBOX_305; 1.
DR PROSITE; PS01264; TBOX_306; 1.
DR PROSITE; PS01264; TBOX_307; 1.
DR PROSITE; PS01264; TBOX_308; 1.
DR PROSITE; PS01264; TBOX_309; 1.
DR PROSITE; PS01264; TBOX_310; 1.
DR PROSITE; PS01264; TBOX_311; 1.
DR PROSITE; PS01264; TBOX_312; 1.
DR PROSITE; PS01264; TBOX_313; 1.
DR PROSITE; PS01264; TBOX_314; 1.
DR PROSITE; PS01264; TBOX_315; 1.
DR PROSITE; PS01264; TBOX_316; 1.
DR PROSITE; PS01264; TBOX_317; 1.
DR PROSITE; PS01264; TBOX_318; 1.
DR PROSITE; PS01264; TBOX_319; 1.
DR PROSITE; PS01264; TBOX_320; 1.
DR PROSITE; PS01264; TBOX_321; 1.
DR PROSITE; PS01264; TBOX_322; 1.
DR PROSITE; PS01264; TBOX_323; 1.
DR PROSITE; PS01264; TBOX_324; 1.
DR PROSITE; PS01264; TBOX_325; 1.
DR PROSITE; PS01264; TBOX_326; 1.
DR PROSITE; PS01264; TBOX_327; 1.
DR PROSITE; PS01264; TBOX_328; 1.
DR PROSITE; PS01264; TBOX_329; 1.
DR PROSITE; PS01264; TBOX_330; 1.
DR PROSITE; PS01264; TBOX_331; 1.
DR PROSITE; PS01264; TBOX_332; 1.
DR PROSITE; PS01264; TBOX_333; 1.
DR PROSITE; PS01264; TBOX_334; 1.
DR PROSITE; PS01264; TBOX_335; 1.
DR PROSITE; PS01264; TBOX_336; 1.
DR PROSITE; PS01264; TBOX_337; 1.
DR PROSITE; PS01264; TBOX_338; 1.
DR PROSITE; PS01264; TBOX_339; 1.
DR PROSITE; PS01264; TBOX_340; 1.
DR PROSITE; PS01264; TBOX_341; 1.
DR PROSITE; PS01264; TBOX_342; 1.
DR PROSITE; PS01264; TBOX_343; 1.
DR PROSITE; PS01264; TBOX_344; 1.
DR PROSITE; PS01264; TBOX_345; 1.
DR PROSITE; PS01264; TBOX_346; 1.
DR PROSITE; PS01264; TBOX_347; 1.
DR PROSITE; PS01264; TBOX_348; 1.
DR PROSITE; PS01264; TBOX_349; 1.
DR PROSITE; PS01264; TBOX_350; 1.
DR PROSITE; PS01264; TBOX_351; 1.
DR PROSITE; PS01264; TBOX_352; 1.
DR PROSITE; PS01264; TBOX_353; 1.
DR PROSITE; PS01264; TBOX_354; 1.
DR PROSITE; PS01264; TBOX_355; 1.
DR PROSITE; PS01264; TBOX_356; 1.
DR PROSITE; PS01264; TBOX_357; 1.
DR PROSITE; PS01264; TBOX_358; 1.
DR PROSITE; PS01264; TBOX_359; 1.
DR PROSITE; PS01264; TBOX_360; 1.
DR PROSITE; PS01264; TBOX_361; 1.
DR PROSITE; PS01264; TBOX_362; 1.
DR PROSITE; PS01264; TBOX_363; 1.
DR PROSITE; PS01264; TBOX_364; 1.
DR PROSITE; PS01264; TBOX_365; 1.
DR PROSITE; PS01264; TBOX_366; 1.
DR PROSITE; PS01264; TBOX_367; 1.
DR PROSITE; PS01264; TBOX_368; 1.
DR PROSITE; PS01264; TBOX_369; 1.
DR PROSITE; PS01264; TBOX_370; 1.
DR PROSITE; PS01264; TBOX_371; 1.
DR PROSITE; PS01264; TBOX_372; 1.
DR PROSITE; PS01264; TBOX_373; 1.
DR PROSITE; PS01264; TBOX_374; 1.
DR PROSITE; PS01264; TBOX_375; 1.
DR PROSITE; PS01264; TBOX_376; 1.
DR PROSITE; PS01264; TBOX_377; 1.
DR PROSITE; PS01264; TBOX_378; 1.
DR PROSITE; PS01264; TBOX_379; 1.
DR PROSITE; PS01264; TBOX_380; 1.
DR PROSITE; PS01264; TBOX_381; 1.
DR PROSITE; PS01264; TBOX_382; 1.
DR PROSITE; PS01264; TBOX_383; 1.
DR PROSITE; PS01264; TBOX_384; 1.
DR PROSITE; PS01264; TBOX_385; 1.
DR PROSITE; PS01264; TBOX_386; 1.
DR PROSITE; PS01264; TBOX_387; 1.
DR PROSITE; PS01264; TBOX_388; 1.
DR PROSITE; PS01264; TBOX_389; 1.
DR PROSITE; PS01264; TBOX_390; 1.
DR PROSITE; PS01264; TBOX_391; 1.
DR PROSITE; PS01264; TBOX_392; 1.
DR PROSITE; PS01264; TBOX_393; 1.
DR PROSITE; PS01264; TBOX_394; 1.
DR PROSITE; PS01264; TBOX_395; 1.
DR PROSITE; PS01264; TBOX_396; 1.
DR PROSITE; PS01264; TBOX_397; 1.
DR PROSITE; PS01264; TBOX_398; 1.
DR PROSITE; PS01264; TBOX_399; 1.
DR PROSITE; PS01264; TBOX_400; 1.
DR PROSITE; PS01264; TBOX_401; 1.
DR PROSITE; PS01264; TBOX_402; 1.
DR PROSITE; PS01264; TBOX_403; 1.
DR PROSITE; PS01264; TBOX_404; 1.
DR PROSITE; PS01264; TBOX_405; 1.
DR PROSITE; PS01264; TBOX_406; 1.
DR PROSITE; PS01264; TBOX_407; 1.
DR PROSITE; PS01264; TBOX_408; 1.
DR PROSITE; PS01264; TBOX_409; 1.
DR PROSITE; PS01264; TBOX_410; 1.
DR PROSITE; PS01264; TBOX_411; 1.
DR PROSITE; PS01264; TBOX_412; 1.
DR PROSITE; PS01264; TBOX_413; 1.
DR PROSITE; PS01264; TBOX_414; 1.
DR PROSITE; PS01264; TBOX_415; 1.
DR PROSITE; PS01264; TBOX_416; 1.
DR PROSITE; PS01264; TBOX_417; 1.
DR PROSITE; PS01264; TBOX_418; 1.
DR PROSITE; PS01264; TBOX_419; 1.
DR PROSITE; PS01264; TBOX_420; 1.
DR PROSITE; PS01264; TBOX_421; 1.
DR PROSITE; PS01264; TBOX_422; 1.
DR PROSITE; PS01264; TBOX_423; 1.
DR PROSITE; PS01264; TBOX_424; 1.
DR PROSITE; PS01264; TBOX_425; 1.
DR PROSITE; PS01264; TBOX_426; 1.
DR PROSITE; PS01264; TBOX_427; 1.
DR PROSITE; PS01264; TBOX_428; 1.
DR PROSITE; PS01264; TBOX_429; 1.
DR PROSITE; PS01264; TBOX_430; 1.
DR PROSITE; PS01264; TBOX_431; 1.
DR PROSITE; PS01264; TBOX_432; 1.
DR PROSITE; PS01264; TBOX_433; 1.
DR PROSITE; PS01264; TBOX_434; 1.
DR PROSITE; PS01264; TBOX_435; 1.
DR PROSITE; PS01264; TBOX_436; 1.
DR PROSITE; PS01264; TBOX_437; 1.
DR PROSITE; PS01264; TBOX_438; 1.
DR PROSITE; PS01264; TBOX_439; 1.
DR PROSITE; PS01264; TBOX_440; 1.
DR PROSITE; PS01264; TBOX_441; 1.
DR PROSITE; PS01264; TBOX_442; 1.
DR PROSITE; PS01264; TBOX_443; 1.
DR PROSITE; PS01264; TBOX_444; 1.
DR PROSITE; PS01264; TBOX_445; 1.
DR PROSITE; PS01264; TBOX_446; 1.
DR PROSITE; PS01264; TBOX_447; 1.
DR PROSITE; PS01264; TBOX_448; 1.
DR PROSITE; PS01264; TBOX_449; 1.
DR PROSITE; PS01264; TBOX_450; 1.
DR PROSITE; PS01264; TBOX_451; 1.
DR PROSITE; PS01264; TBOX_452; 1.
DR PROSITE; PS01264; TBOX_453; 1.
DR PROSITE; PS01264; TBOX_454; 1.
DR PROSITE; PS01264; TBOX_455; 1.
DR PROSITE; PS01264; TBOX_456; 1.
DR PROSITE; PS01264; TBOX_457; 1.
DR PROSITE; PS01264; TBOX_458; 1.
DR PROSITE; PS01264; TBOX_459; 1.
DR PROSITE; PS01264; TBOX_460; 1.
DR PROSITE; PS01264; TBOX_461; 1.
DR PROSITE; PS01264; TBOX_462; 1.
DR PROSITE; PS01264; TBOX_463; 1.
DR PROSITE; PS01264; TBOX_464; 1.
DR PROSITE; PS01264; TBOX_465; 1.
DR PROSITE; PS01264; TBOX_466; 1.
DR PROSITE; PS01264; TBOX_467; 1.
DR PROSITE; PS01264; TBOX_468; 1.
DR PROSITE; PS01264; TBOX_469; 1.
DR PROSITE; PS01264; TBOX_470; 1.
DR PROSITE; PS01264; TBOX_471; 1.
DR PROSITE; PS0
```

Wed Oct 9 10:30:06 2002

```

OC 10 LAARAGGKGGGIEGP 26
OC II: ||||| ||||| ||
OC 502 LASVAGGNGGGG--GP 516
OX NCBI_TaxID=3702;
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA; TISSUE=Leaf;
RC MEDLINE=98090460; PubMed=9427751;
RX Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Saizberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
CC -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC STAGES.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U91995; AAC18440.1;
CC EMBL; AC007932; AAD49755.1;
CC Mendel; 24973; Arath; 3192;24973.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC Developmental protein.
FT DOMAIN 13 104 GLY-RICH.
FT SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;
SQ
Query Match 28.0%; Score 54; DB 1; Length 1048;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 15 GGGKGGGGTGTQ 30
DB 98 GGGGGGGGGSSGPPQ 113
RESULT 22
HS70_SCHJA

```

ID HS70_SCHJA STANDARD; PRT; 198 AA.
 AC P12795;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last annotation update)
 DE Heat shock 70 kDa protein (HSP70) (Fragment).
 OS Schistosoma japonicum (Blood fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 CC Rhabditophora; Eulicthophora; Revertospermata; Mediofusata;
 CC Neodermata; Trematoda; Digenea; Strigeidae; Schistosomatoidea;
 CC Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6182;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88318804; PubMed=2457805;
 RA Hedstrom R., Culpepper J., Schinski V., Agabian N., Newport G.;
 RT "Schistosome heat-shock proteins are immunologically distinct
 RT host-like antigens."
 RL Mol. Biochem. Parasitol. 29:275-282(1988).
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL; M21011; AAA29897.1; -
 DR PIR; A54507; A54507.
 DR HSP; P08109; ICKR.
 DR InterPro; IPR001023; HSP70.
 DR PROSITE; PS00297; HSP70.1; PARTIAL.
 DR PROSITE; PS00329; HSP70.2; PARTIAL.
 DR PROSITE; PS01036; HSP70.3; PARTIAL.
 KW ATP-binding; Heat shock.
 FT NON_TER
 SQ SEQUENCE 198 AA; 21845 MW; 800F8596046D5313 CRC64;
 Query Match 27.7%; Score 53.5; DB 1; Length 198;
 Best Local Similarity 41.4%; Pred. No. 28;
 Matches 12; Conservative 3; Mismatches 3; Indels 11; Gaps 1;
 Qy 13 RAGG-----GKGGGGTEGPTLRQ 30
 Db 168 RAGGVPSPGMPGMPGAGGGGKGPTIE 196
 RESULT 23
 PAC4_HUMAN STANDARD; PRT; 969 AA.
 ID PAC4_HUMAN
 AC F29122; Q15099; Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
 AC Q9UEG7; Q9Y4G9; Q9Y4H0; Q9Y4H1;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like proprotein
 DE convertase 4) (SPC4).
 GN PACE4.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).
 RC TISSUE=Hepatoma, and Kidney;
 RX MEDLINE=92075167; PubMed=1741956;
 RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
 RA Barr P.J.;
 RT "Identification of a second human subtilisin-like protease gene in
 RT the fes/fps region of chromosome 15."

RL DNA Cell Biol. 10:757-769(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).
 RC TISSUE=Placenta;
 RX MEDLINE=94235049; PubMed=8179631;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms."
 RL Biochem. Biophys. Res. Commun. 200:943-950(1994).
 RN [3]
 RP ERRATUM.
 RX MEDLINE=95071480; PubMed=7980617;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms."
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM PACE4A-II).
 RC TISSUE=Placenta;
 RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
 RA Matsuda Y.;
 RT "Identification of a novel PACE4 isoform, PACE4E."
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).
 RC TISSUE=Cerebellum;
 RX MEDLINE=97335942; PubMed=9192737;
 RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
 RA Akamatsu T., Nagamune H., Matsuda Y.;
 RT "A novel human PACE4 isoform, PACE4E is an active processing protease
 RT containing a hydrophobic cluster at the carboxy terminus."
 RL J. Biochem. 121:941-948(1997).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; A-II; CS; D; E-I; E-II).
 RX MEDLINE=98021085; PubMed=9378725;
 RA Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
 RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
 RT "Genomic organization and alternative splicing of human PACE4 (SPC4),
 RT kexin-like processing endoprotease."
 RL J. Biochem. 122:438-452(1997).
 RN [7]
 RP ALTERNATIVE SPLICING (ISOFORM PACE4CS).
 RX MEDLINE=97064242; PubMed=8906861;
 RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
 RT "Functional analysis of human PACE4-A and PACE4-C isoforms;
 RT identification of a new PACE4-CS isoform."
 RL FEBS Lett. 396:31-36(1996).
 RN [8]
 RP CHARACTERIZATION.
 RX MEDLINE=99233559; PubMed=10215603;
 RA Sucic J.F., Moehring J.N., Inocencio N.M., Luchini J.W.,
 RA Moehring T.J.;
 RT "Endoprotease PACE4 is Ca2+-dependent and temperature-sensitive and
 RT can partly rescue the phenotype of a furin-deficient cell strain."
 RL Biochem. J. 339:639-647(1999).
 RN [9]
 RP PROCESSING.
 RX MEDLINE=98408849; PubMed=9738469;
 RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
 RA Tsuji A., Matsuda Y.;
 RT "Biosynthetic processing and quaternary interactions of proprotein
 RT convertase SPC4 (PACE4)."
 RL FEBS Lett. 434:155-159(1998).
 CC -!- FUNCTION: LIKELY TO REPRESENT AN ENDOPEPTIDASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS,
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -!- COFACTOR: PACE4A IS PROBABLY CALCIUM-DEPENDENT.

-!- SUBUNIT: THE PACE4A-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX WHEREAS NATURE PACE4A-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.

-!- SUBCELLULAR LOCATION: PACE4A-I AND PACE4A-II ARE SECRETED. PACE4C AND PACE4CS ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-TERMINUS. PACE4B MIGHT BE SECRETED.

-!- ALTERNATIVE PRODUCTS: 8 ISOFORMS: PACE4A-I/PACE4 (SHOWN HERE), PACE4A-II, PACE4B/PACE4.1, PACE4C, PACE4CS, PACE4D, PACE4E-I AND PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B, C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.

-!- TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT COMPARATIVELY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST COMPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE EXPRESSED IN PLACENTA. PACE4E-I IS EXPRESSED IN CEREBELLUM, PLACENTA AND PITUITARY. PACE4E-II IS AT LEAST PRESENT IN CEREBELLUM.

-!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC RETICULUM. ISOFORM PACE4D LACKS THE PROPEPTIDE DOMAIN.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE SUBTILASE FAMILY.

-!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

Query Match 27.7%; Score 53.5; DB 1; Length 969;
Best Local Similarity 44.8%; Pred. No. 1.1e+02;
Matches 13; Conservative 1; Mismatches 8; Indels 7; Gaps 1;

11 AARAGGGKGGGIEGPTLR-----QWL 32

RESULT 24

STANDARD:	PRT:	332 AA:
YACO_ALCEU		
YACO_ALCEU		

P31640; 01-JUL-1993 (Rel. 26, Created)
AD 01-JUL-1993 (Rel. 26, Last sequence update)
ID 01-JUL-1993 (Rel. 26, Last annotation update)
DC 16-OCT-2001 (Rel. 40, Last annotation update)
DT Hypothetical protein in acO3' region (ORF2) (Fragment).
TT Hypothetical protein in acO3' region (ORF2) (Fragment).
AC Alcaligenes eutrophus (Ralstonia eutropha).
ID Ralstonia eutropha (Alcaligenes eutrophus).
DC beta subdivision; Ralstonia group;
OS Proteobacteria.
OT

OC Ralstonia.
OX NCBI TaxID=510;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=93015711; PubMed=13565

RA Priefert H., Steinbuechel A.;

RT "Identification and molecular characterization of

RT A synthetase gene (acoE) of *Alcaligenes eutrophus*.¹

RL J. Bacteriol. 174:6590-6599(1992).

-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (LARGE)
CC I- SPECIFICITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).

-I- SIMILARITY: BELONGS TO THE SODIUM-COCTATE DASH GROUP

 this swiss-ppom entry is copyright. It is produced through a collabora

CC THIS SWISS-PROT entry is copyrighted by the EMBL outsta
CC between the Swiss Institute of Bioinformatics and the EMBL outsta

CC Between the Swiss Institute and the European Bioinformatics Institute. There are no restrictions on

use by non-profit institutions as long as its content is in the public domain.

CC modified and this statement is not removed. Usage by and for comm

CC entities requires a license agreement (See <http://www.1sb-slb.ch/ann>

CC or send an email to license@isb-sib.ch).

22

DR EMBL; M97217; AAA21946.1; -.

[REDACTED]

100

DR PIR; B45736; B45736.
 DR InterPro; IPR001734; Na_solut_sympo1.
 DR PROSITE; PS00456; NA_SOLUT_SYMP_1; PARTIAL.
 DR PROSITE; PS00457; NA_SOLUT_SYMP_2; PARTIAL.
 DR PROSITE; PS00283; NA_SOLUT_SYMP_3; 1.
 KW Hypothetical protein; Transport; Transmembrane; Sodium transport;
 FT Symptot. 332 332
 FT NON_TER 332 332
 SQ SEQUENCE 332 AA; 37031 MW; CB8F4471BD7341C8 CRC64;
 Query Match 27.5%; Score 53; DB 1; Length 332;
 Best Local Similarity 52.4%; Pred. No. 50;
 Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 3 GPTLRQLAARAGGGGGGGI 23
 DB 137 GYIPDELAARYGNGKPGNL 157
 ID SIX3_MOUSE STANDARD; PRT; 333 AA.
 AC Q62233; P70176; P70177;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Homeobox protein SIX3 (Sine oculis homeobox homolog 3).
 GN SIX3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Embryonic brain;
 RX MEDLINE=96125147; PubMed=8575305;
 RA Oliver G., Mailhos A., Wehr R., Copeland N.G., Jenkins N.A.,
 RA Gruss P.;
 RT "Six3, a murine homologue of the sine oculis gene, demarcates the
 RT most anterior border of the developing neural plate and is expressed
 RT during eye development."
 RL Development 121:4045-4055(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RX MEDLINE=96409319; PubMed=8814301;
 RA Kawakami K., Ohto H., Takizawa T., Saito T.;
 RA "Identification and expression of Six family genes in mouse retina";
 RL FEBS Lett. 393:259-263(1996).
 CC -1- FUNCTION: MAY BE INVOLVED IN VISUAL SYSTEM DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIX3A AND SIX3B (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT E6.5 OF EMBRYO DEVELOPMENT
 CC AROUND THE ANTERIOR BORDER. AT E8.5, EXPRESSION IS FOUND OVER THE
 CC ANTERIOR NEURAL PLATE. AT E9.5, IN THE DIENCEPHALIC PART OF THE
 CC VENTRAL FOREBRAIN, OPTIC VESICLES, OLFACTORY PLACODES AND RATHKE'S
 CC POUCH. IN LATER STAGES, PRESENT IN HYPOTHALAMUS, EYES AND
 CC PITUITARY.
 CC -1- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEODOMAIN
 CC PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X90871; CAA62379.1; ALT_INIT.
 CC EMBL; D83144; BAA11822.1; -.
 CC EMBL; D83145; BAA11823.1; -.

DR HSHP; P40427; IB81.
 DR TRANSFAC; T03263; -.
 DR TRANSFAC; T03270; -.
 DR MGD; MGI:102764; Six3.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;
 Alternative splicing.
 FT DOMAIN 33 70
 FT DNA_BIND 207 266
 FT DOMAIN 264 267
 FT VARSPLIC 271 286
 FT VARSPLIC 287 333
 FT CONFLICT 44 44
 FT CONFLICT 118 119
 FT CONFLICT 278 333
 FT FT
 FT FT
 SQ SEQUENCE 333 AA; 35592 MW; 1AD7D3C4388043B9 CRC64;
 Query Match 27.5%; Score 53; DB 1; Length 333;
 Best Local Similarity 62.5%; Pred. No. 50;
 Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 11 AARAGGGKGGGIEGP 26
 DB 59 AGGAGGGGGGGRAP 74
 ID FXD3_CHICK STANDARD; PRT; 394 AA.
 AC P79772;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Forkhead box protein D3 (HNF3/FH transcription factor genesis) (Winged
 DE helix protein CWH-3).
 GN FOXD3.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=97141794; PubMed=8988052;
 RA Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;
 RT "Aberrant cell growth induced by avian winged helix proteins";
 RL Cancer Res. 57:123-129(1997).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U37274; AAC60066.1; -.
 DR HSHP; Q63245; 2HFH.
 DR TRANSFAC; T02495; -.
 DR InterPro; IPR001766; Fork_head.
 DR Pfam; PF00250; Fork_head; 1.

Matches 11; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

```

DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
DR FT DOMAIN 67 70 POLY-ALA.
DR FT DOMAIN 80 91 POLY-GLY.
DR FT DOMAIN 100 106 POLY-ALA.
DR FT DNA_BIND 117 211 FORK-HEAD.
DR SQ SEQUENCE 394 AA; 40995 MW; 3244AB36B9E31899 CRC64;

Query Match 27.5%; Score 53; DB 1; Length 394;
Best Local Similarity 76.9%; Pred. No. 58;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 13 RAGGKGKGGGIEG 25
Db 82 RGGGGGGGGEG 94

RESULT 27
HKLBL_LYCES STANDARD; PRT; 426 AA.
AC 022300;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Homeobox protein knotted-1 like LET12.
GN LET12.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98145476; PubMed=9484482;
RX Janssen B.J., Williams A., Chen J.J., Mathern J., Hake S., Sinha N.;
RT "Isolation and characterization of two knotted-like homeobox genes
from tomato.";
RL Plant Mol. Biol. 36:417-425(1998).
CC -!- FUNCTION: MAY HAVE A ROLE TO PLAY IN FORMATIVE EVENTS IN OVULE AND
EMBRYO MORPHOGENESIS.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN THE MATURE PLANT.
CC -!- SIMILARITY: BELONGS TO THE TALE/KNOX FAMILY OF HOMEBOX PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF000142; AAC49918.1; -.
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
DR FT DOMAIN 15 24 POLY-GLN.
DR FT DOMAIN 69 76 POLY-ALA.
DR FT DOMAIN 140 152 POLY-ASN.
DR FT DOMAIN 283 287 POLY-ASP.
DR FT DOMAIN 325 348 ELK DOMAIN.
DR FT DNA_BIND 349 411 HOMEBOX (TALE-TYPE).
DR SQ SEQUENCE 426 AA; 47581 MW; 5B52B9E0A34A86BC CRC64;

Query Match 27.5%; Score 53; DB 1; Length 426;
Best Local Similarity 64.7%; Pred. No. 62;

Matches 11; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

Oy 8 OWLA--ARAGGKGGGG 22
Db 96 QWLSPTAAAGGGGNGG 112

RESULT 28
OC3N_HUMAN STANDARD; PRT; 443 AA.
AC P20265; Q14960;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nervous-system specific octamer-binding transcription factor N-OCT 3
DE (Brain-specific homeobox/POU domain protein 2) (BRN-2 protein)
DE [Contains: N-OCT 5A; N-OCT 5B].
GN POU3F2 OR BRN2 OR OTF7 OR OCT7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93181199; PubMed=8441633;
RA Schreiber E., Tobler A., Malpiero U., Schaffner W., Fontana A.;
RT "cDNA cloning of human N-Oct3, a nervous-system specific POU domain
transcription factor binding to the octamer DNA motif.";
RL Nucleic Acids Res. 21:253-258(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95380176; PubMed=7651733;
RA Angus J., Thomson F., Murphy K., Baker E., Sutherland G.R.,
RA Parsons P.G., Sturm R.A.;
RT "The brn-2 gene regulates the melanocytic phenotype and tumorigenic
potential of human melanoma cells.";
RL Oncogene 11:691-700(1995).
RN [3]
RP SEQUENCE OF 280-404 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89295573; PubMed=2739723;
RA He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,
RA Rosenfeld M.G.;
RT "Expression of a large family of POU-domain regulatory genes in
mammalian brain development.";
RL Nature 340:35-42(1989).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
OF 0, 2 OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
PROMOTERS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; N-OCT 3 (SHOWN HERE), N-OCT 5A
AND N-OCT 5B; ARE PRODUCED BY ALTERNATIVE INITIATION.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
CELL LINEAGE.
CC -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
TO CLASS-3 POU.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z11933; CAA77990.1; -.
DR EMBL; L37868; AAB59611.1; -.
DR PIR; S05043; S05043.

```


Wed Oct 9 10:30:06 2002

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED

CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-

CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE

CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH

CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP

CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.

CC -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AL021287; CAAL6096.1; -;

DR EMBL; AE007128; AAK47420.1; -;

DR TIGR; MT3091; -;

DR TubercuList; Rv3011c; -;

DR InterPro; IPR000120; Amidase.

DR Pfam; PF01425; Amidase; 1.

DR PROSITE; PS00571; AMIDASES; 1.

KW Protein biosynthesis; Ligase; Complete proteome.

FT CONFLICT 420 420 M -> L (IN REF. 2).

SQ SEQUENCE 494 AA; 51438 MW; 99A8824ABC436CA6 CRC64;

Query Match 27.5%; Score 53; DB 1; Length 494;

Best Local Similarity 52.6%; Pred No. 71;

Matches 10; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 GPTLRQWLAAAGGKGKG 21

Db 141 GPTRNPNLDRVPGSGGG 159

Search completed: October 9, 2002, 09:00:18
 Job time : 6.3831 secs

```

OC Rhabdittidae; Peloderinae; Caenorhabdittis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92375684; PubMed-1354852;
RA Iwasaki M., Okumura K., Kondo Y., Igarashi H., Tanaka T.;
RT "cDNA cloning of a novel heterogeneous nuclear ribonucleoprotein gene
RT homologue in Caenorhabdittis elegans using hamster prion protein cDNA
RL as a hybridization probe.";
RL Nucleic Acids Res. 20:4001-4007(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Du Z., Scheet P., Andrews S.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; D10877; BAA01645.1; -;
DR EMBL; AF038613; AAB92051.1; -;
DR HSSP; P09651; IUP1.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000504; RRM.
DR PRINTS; PR01228; itm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 2.
KW RNA-binding.
SQ SEQUENCE 346 AA; 36344 MW; 48B95818DBB9A54 CRC64;

Query Match 33.3%; Score 74.5; DB 5; Length 346;
Best Local Similarity 43.6%; Pred. No. 2.5;
Matches 17; Conservative 3; Mismatches 16; Indels 3; Gaps 1;

QY 1 GGGGGIEGPTLRQ---WLAARAGGGGGGGIEGPTLRQW 36
  |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 289 GGCGGGGPGQQQGGGPGQQGGGGGGGGGGGGGGG 327

RESULT 30
O02402
ID O02402 PRELIMINARY; PRT; 738 AA.
AC O02402;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE INSOLUBLE PROTEIN.
OS Pinctada fucata.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;
OC Pterioidea; Pteriidae; Pinctada.
OX NCBI_TaxID=50426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97320490; PubMed-9177341;
RA Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,
RA Nakashima K., Takahashi T.;
RT "Structures of mollusc shell framework proteins.";
RL Nature 387:563-564(1997).
DR EMBL; D86074; BAA20466.1; -;
SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

Query Match 33.3%; Score 74.5; DB 5; Length 738;
Best Local Similarity 55.0%; Pred. No. 5.4;
Matches 22; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

QY 1 GGGGGIEGPTLRQWLAAAGGG-GGGGGIEGPTLRQWLAA 39
  |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 450 GGGGGGAGALAAALAAAGAGGGGGLGGGG--GGALAAALAA 487

```

Search completed: October 9, 2002, 09:03:18
 Job time : 16.7742 secs

```

SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=107311132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Butcher J., Brockstein P., Bottier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphree L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
RA Svirskaas R., Tector C., Turner R., Venter E., Weissbach J.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Wu D., Yang S., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong G.M., Weinstock G.M., Weissbach J.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003459; AAC46950.1; -.
DR HSSP; Q58108; 1PEN.
DR Flybase; FBgn003062; Fib.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR00692; Fibrillarlin.
DR Pfam; PF01269; Fibrillarlin; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR00052; FIBRILLARIN.
DR PRODOM; PD004637; Fibrillarlin; 1.
DR PROSITE; PS00566; FIBRILLARIN; 1.
SQ SEQUENCE 344 AA; 34637 MW; 58B536FAACAE01D6 CRC64;

Query Match 33.3%; Score 74.5; DB 5; Length 344;
Best Local Similarity 55.6%; Pred. No. 2.5;
Matches 15; Conservative 0; Mismatches 3; Indels 9; Gaps 1;

QY 1 GGGGIEGTPLRWLAARAGGGGGG 27
|| || || : | | | | | | | |
Db 24 GGGGFRG-----RGGGGGGG 41

RESULT 29
Q22037 PRELIMINARY; PRT; 346 AA.
ID Q22037
AC Q22037
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HNRNP LIKE PROTEIN.
GN RBP-1 OR F42A6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
RN [1]

SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=107311132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Butcher J., Brockstein P., Bottier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphree L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
RA Svirskaas R., Tector C., Turner R., Venter E., Weissbach J.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Wu D., Yang S., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong G.M., Weinstock G.M., Weissbach J.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003459; AAC46950.1; -.
DR HSSP; Q58108; 1PEN.
DR Flybase; FBgn003062; Fib.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR00692; Fibrillarlin.
DR Pfam; PF01269; Fibrillarlin; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR00052; FIBRILLARIN.
DR PRODOM; PD004637; Fibrillarlin; 1.
DR PROSITE; PS00566; FIBRILLARIN; 1.
SQ SEQUENCE 344 AA; 34637 MW; 58B536FAACAE01D6 CRC64;

Query Match 33.3%; Score 74.5; DB 5; Length 308;
Best Local Similarity 43.6%; Pred. No. 2.3;
Matches 17; Conservative 3; Mismatches 16; Indels 3; Gaps 1;

QY 1 GGGGIEGTPLRWLAARAGGGGGG 27
|| || || : | | | | | | | |
Db 251 GGGGGGPPQQQQGGGPGQQGGGGGGGGGGQG 289

RESULT 28
Q9WIV3 PRELIMINARY; PRT; 344 AA.
ID Q9WIV3
AC Q9WIV3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE FIB PROTEIN.
GN FIB OR CG9888.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyrdoidea; Drosophilidae; Drosophila.
OS NCBI_TaxID=7227;
RN [1]

```


Wed Oct 10 10:30:21 2002

```

OS Lampetra fluviatilis (River lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Lampetra.
OX NCBI_TaxID=7748;
RN [1]
RP SEQUENCE FROM N.A.
RA Kao H.-T., Porton B., Hilfiker S., Stefani G., Pieribone V.A.,
RA Desalle R., Greengard P.;
RT "Molecular Evolution of the Synapsin Gene Family.";
RL J. Exp. Zool. 0:0-0(2000)
DR EMBL; AF192747; AAF08805.1; -
DR HSSP; P17599; IAOX.
DR InterPro; IPR001359; Synapsin.
DR Pfam; PF02078; Synapsin_C; 1.
DR Pfam; PF02750; Synapsin_C; 1.
DR PRINTS; PR01368; SYNAPSIN.
DR PROSITE; PS00415; SYNAPSIN.1; 1.
SQ SEQUENCE 687 AA; 71320 MW; D2D917A69FBDABAC CRC64;

Query Match 33.7%; Score 75.5; DB 13; Length 687;
Best Local Similarity 45.7%; Pred. No. 4;
Matches 16; Conservative 2; Mismatches 6; Indels 11; Gaps 1;

QY 1 GGGGIEGPTLRQWLAARAGGGGGGIEGPTLRQ 35
DB 574 GGGGG-----PRGGSGGGGMVGPOTQ 597

RESULT 21
Q24917 PRELIMINARY; PRT; 250 AA.
ID Q24917
AC Q24917; 1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BINDIN (FRAGMENT).
OS Echinometa sp.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidea; Echinacea; Echinoida; Echinometridae;
OC Echinometra.
OX NCBI_TaxID=43350;
RN [1]
RP SEQUENCE FROM N.A.
RA Metz E.C., Palumbi S.R.;
RX MEDLINE-96164534; PubMed=8587504;
RA Metz E.C., Palumbi S.R.;
RT "Positive selection and sequence rearrangements generate extensive
polymorphism in the gamete recognition protein bindin.";
RL Mol. Biol. Evol. 13:397-406(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Metz E.C., Palumbi S.R.;
RX Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39514; AAB00415.1; -
DR InterPro; IPR000775; Bindin.
DR Pfam; PF02084; Bindin; 1.
DR PRINTS; PR00761; BINDIN.
DR NON_TER 1
FT SEQUENCE 250 AA; 24156 MW; 77FF0F5B8F44D248 CRC64;

Query Match 33.5%; Score 75; DB 5; Length 250;
Best Local Similarity 53.3%; Pred. No. 1.6;
Matches 16; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 1 GGGGIEGPTLRQWLAARAGGGGGGIEG 30
DB 207 GGGGGAGG-----MGWAGRGGGGGGGGMMG 232

RESULT 22
Q24917 PRELIMINARY; PRT; 453 AA.
ID Q24917
AC Q24917; 1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BINDIN (FRAGMENT).
OS Echinometa sp.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidea; Echinacea; Echinoida; Echinometridae;
OC Echinometra.
OX NCBI_TaxID=43350;
RN [1]
RP SEQUENCE FROM N.A.
RA Metz E.C., Palumbi S.R.;
RX MEDLINE-96164534; PubMed=8587504;
RA Metz E.C., Palumbi S.R.;
RT "Positive selection and sequence rearrangements generate extensive
polymorphism in the gamete recognition protein bindin.";
RL Mol. Biol. Evol. 13:397-406(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Metz E.C., Palumbi S.R.;
RX Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39514; AAB00415.1; -
DR InterPro; IPR000775; Bindin.
DR Pfam; PF02084; Bindin; 1.
DR PRINTS; PR00761; BINDIN.
DR NON_TER 1
FT SEQUENCE 250 AA; 24156 MW; 77FF0F5B8F44D248 CRC64;

Query Match 33.5%; Score 75; DB 5; Length 250;
Best Local Similarity 53.3%; Pred. No. 1.6;
Matches 16; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 1 GGGGIEGPTLRQWLAARAGGGGGGIEG 30
DB 207 GGGGGAGG-----MGWAGRGGGGGGGGMMG 232

RESULT 23
Q24917 PRELIMINARY; PRT; 904 AA.
ID Q24917
AC Q24917; 1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NONGRADIANT BYSSAL.
OS Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RA Metz E.C., Palumbi S.R.;
RX MEDLINE-98393676; PubMed=9724735;
RA Qin X.X., Waite J.H.;
RT "A potential mediator of collagenous block copolymer gradients in
mussel byssal threads.";
RL proc. Natl. Acad. Sci. U.S.A. 95:10517-10522(1998).
DR EMBL; AF043944; AAC33847.1; -
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 7.
DR SEQUENCE 904 AA; 77883 MW; 5529135651AD4C40 CRC64;

Query Match 33.5%; Score 75; DB 5; Length 904;
Best Local Similarity 55.6%; Pred. No. 5.8;
Matches 15; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 2 GGGGIEGPTLRQWLAARAGGGGGGGI 28
DB 157 GGGGIEGPTLRQWLAARAGGGGGGGI 183

```

```

AC Q9N6M8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OVO PROTEIN (FRAGMENT).
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RA Begun D.J., Whitley P.;
RX MEDLINE-20283933; PubMed=10823947;
RA "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252749; AAF68532.1; -
DR EMBL; AF252744; AAF68527.1; -
DR EMBL; AF252745; AAF68528.1; -
DR EMBL; AF252746; AAF68529.1; -
DR EMBL; AF252747; AAF68530.1; -
DR EMBL; AF252748; AAF68531.1; -
DR Flybase; FBgn0014829; DsimOvo.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHELL.
DR NON_TER 1
FT SEQUENCE 453 AA; 46322 MW; 32677BA0F49648FC CRC64;

Query Match 33.5%; Score 75; DB 5; Length 453;
Best Local Similarity 55.2%; Pred. No. 3;
Matches 16; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 2 GGGGIEGPTLRQWLAARAGGGGGGIEG 30
DB 216 GGGGAGP--GGGPSANSGGGGGGGXNG 242

RESULT 23
Q24917 PRELIMINARY; PRT; 904 AA.
ID Q24917
AC Q24917; 1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NONGRADIANT BYSSAL.
OS Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RA Metz E.C., Palumbi S.R.;
RX MEDLINE-98393676; PubMed=9724735;
RA Qin X.X., Waite J.H.;
RT "A potential mediator of collagenous block copolymer gradients in
mussel byssal threads.";
RL proc. Natl. Acad. Sci. U.S.A. 95:10517-10522(1998).
DR EMBL; AF043944; AAC33847.1; -
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 7.
DR SEQUENCE 904 AA; 77883 MW; 5529135651AD4C40 CRC64;

Query Match 33.5%; Score 75; DB 5; Length 904;
Best Local Similarity 55.6%; Pred. No. 5.8;
Matches 15; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 2 GGGGIEGPTLRQWLAARAGGGGGGGI 28
DB 157 GGGGIEGPTLRQWLAARAGGGGGGGI 183

```


RP SEQUENCE FROM N.A.
RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,
RL Schueller C.;
RN Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022140; CAA18105.1; -;
DR EMBL; AL161556; CAB79157.1; -;
DR InterPro; IPR002173; PfkB.
DR PROSITE; PS00583; PFKB_KINASES.1; UNKNOWN.1.
SQ SEQUENCE 396 AA; 31470 MW; 1CDDDOFFBE337CAB CRC64;

Query Match 33.7%; Score 75.5; DB 10; Length 396;
Best Local Similarity 53.3%; Pred. No. 2.3;
Matches 16; Conservative 1; Mismatches 6; Indels 7; Gaps 1;

QY 1 GGGGGIEGPTLRQWLAAARAGGGGGGIEG 30
Db 239 GAGGGVSG-----AAGGGGGGGGGSG 261
||||| | || ||||| |

RESULT 19
O9PUD8 PRELIMINARY; PRT; 642 AA.
ID Q9PUD8
AC Q9PUD8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
GN SYN I.
OS Lampetra fluviatilis (River lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Lampetra.
OX NCBI_TaxID=7748;
RN [1]
RP SEQUENCE FROM N.A.
RA Kao H.-T., Porton B., Hilfiker S., Stefani G., Pieribone V.A.,
RA Desalle R., Greengard P.;
RT "Molecular Evolution of the Synapsin Gene Family.";
RL J. Exp. Zool. 0:0-0(2000).
DR EMBL; AF192748; AAF08806.1; -;
DR HSSP; PI7599; LAUX.
DR InterPro; IPR001359; Synapsin.
DR Pfam; PF02078; Synapsin; 2.
DR Pfam; PF02750; Synapsin C; 1.
DR PRINTS; PR01368; SYNAPSIN.
DR PROSITE; PS00415; SYNAPSIN.1.
SQ SEQUENCE 642 AA; 66448 MW; BE3FDB68A3A89CB8 CRC64;

Query Match 33.7%; Score 75.5; DB 13; Length 642;
Best Local Similarity 45.7%; Pred. No. 3.7;
Matches 16; Conservative 2; Mismatches 6; Indels 11; Gaps 1;

QY 1 GGGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQ 35
Db 574 GGGGG-----PRPGGGGGGGVMGPQTQ 597
||||| | || ||||| |

RESULT 20
O9PUD9 PRELIMINARY; PRT; 687 AA.
ID Q9PUD9
AC Q9PUD9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN SYN I.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

Wed Oct '9 10:30:21 2002

```
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U40802; AAK19014.2; -.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHELL.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 22384 MW; EFE679D8E0AE216 CRC64;

Query Match 34.4%; Score 77; DB 5; Length 265;
Best Local Similarity 63.0%; Pred. No. 1.1;
Matches 17; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 1 GGGGGIEGPTLRQWLAARAGGGGGGGG 27
   ||||| | | | | | | | | | | |
Db 233 GGGGGIPG-----QSVYMGAGGGGGGGG 255

RESULT 14
Q9W3E3 PRELIMINARY; PRT; 679 AA.
AC Q9W3E3;
DT 01-MAY-2000 (TREMELREL. 13, Created)
DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE CG1521 PROTEIN.
GN CG1521.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Benos P.V., Bertram B.P., Brokstein P., Brotter P.,
RA Beeson K.Y., Bock J., Bock J., Bock J., Bock J., Bock J., Bock J.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Center A., Chandra I.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Cadenhead L., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,
RA Spier E., Spradling A.C., Stapleton M., Turner E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Zhao L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
```

```

DR Pfam: PF00932; IF_tail; 1.
SQ SEQUENCE 650 AA; 70492 MW; 3F45D74AAC220CE9 CRC64;

Query Match      35.3%; Score 79; DB 5; Length 650;
Best Local Similarity 46.5%; Pred. No. 1.7;
Matches 20; Conservative 7; Mismatches 10; Indels 6; Gaps 2;

QY 1 GGGGIEGPTLRQWLAARAGGGG-----GGGIEGPTLRQWLA 39
Db 42 GGGG--GMSITSMSRSGGGGFGMGGGGMSLKMVRVA 82

RESULT 10
Q943K0
AC Q943K0 PRELIMINARY; PRT; 253 AA.
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE P0039A07.6 PROTEIN.
GN P0039A07.6
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0039A07."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003235; BAB64100.1;
SQ SEQUENCE 253 AA; 25568 MW; A963166CE5F97B2B CRC64;

Query Match      34.8%; Score 78; DB 10; Length 253;
Best Local Similarity 35.0%; Pred. No. 0.82;
Matches 21; Conservative 3; Mismatches 10; Indels 26; Gaps 2;

QY 1 GGGGGIEG-----PTLRQW-----LAARAGGGGGGIEGPTLR 34
Db 47 GGGGGGGGRRASVGVGPTREFSGRNGCRVGTGVYRVAYRAGAGGGGGGPRGFALK 106

RESULT 11
Q9LG49
AC Q9LG49 PRELIMINARY; PRT; 308 AA.
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE ESTS AU029606(E31139).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0699D11."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002817; BAB03442.1;
DR InterPro; IPR003015; HLH_MYC.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOW.1.
SQ SEQUENCE 308 AA; 33320 MW; 4A3A3B83E6DE468 CRC64;

Query Match      34.8%; Score 78; DB 10; Length 308;
Best Local Similarity 54.8%; Pred. No. 1;
Matches 17; Conservative 3; Mismatches 5; Indels 6; Gaps 2;

```

```

QY 1 GGGGI-EGPTLRQWLAARAGGGGGGIEG 30
      |||||  ||  |  |||||  ||  |
Db 8 GGGGVNAGPGV-----AGGGGGGGGGVG 33

```

```

RESULT 12
O35295
ID O35295 PRELIMINARY; PRT; 324 AA.
AC O35295;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE VASCULAR ACTIN SINGLE-STRANDED DNA-BINDING FACTOR 2 P44
COMPONENT.
GN PURB OR PURB BETA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J; TISSUE=LUNG;
RX MEDLINE=97476282; PubMed=9334258;
RA Kelm R.J. Jr., Elder P.K., Strauch A.R., Getz M.J.;
RT "Sequence of cDNAs encoding components of vascular actin single-
RT stranded DNA-binding factor 2 establish identity to Puralpha and
RT Purbeta."
RL J. Biol. Chem. 272:26727-26733(1997).
DR EMBL; AF017630; AAB71859.1;
KW MGD; MGI:1338779; Purb.
KW DNA-binding.
SQ SEQUENCE 324 AA; 33901 MW; 5CB70CC3FDB7913 CRC64;

```

```

Query Match      34.8%; Score 78; DB 11; Length 324;
Best Local Similarity 52.5%; Pred. No. 1.1;
Matches 21; Conservative 2; Mismatches 15; Indels 2; Gaps 1;

QY 1 GGGGIEGPTLRQWLAARAGGGGGGIEGPTLRQWLAAR 40
      |||||  ||  |  |||||  ||  |
Db 12 GGGGGGGGGGPGFQ--PAPRGGGGGGGGPGGEQTOELASK 49

```

```

RESULT 13
Q23347
ID Q23347 PRELIMINARY; PRT; 265 AA.
AC Q23347;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ZC477.1.
GN ZC477.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z.;
RT "The sequence of C. elegans cosmid ZC477."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;

```

Q

SEQUENCE 422 AA; 44892 MW; 85FE742F07751B24 CRC64;

Query Match 36.2%; Score 81; DB 5; Length 422;
Best Local Similarity 36.7%; Pred. No. 0.68;
Matches 22; Conservative 3; Mismatches 9; Indels 26; Gaps 2;

y 1 GGCGGEGPTLROWLAARA-----GGGGGGGGIEG-----PTLR 34
||||| | : ||||
53 GGCGRGGASMLVPAGRASAFSSSASFSGGGGGGGGGIISGMWTEKPTMR 112
||||| | : ||||
RESULT 6
ID Q9PVG9 PRELIMINARY; PRT; 431 AA.
AC Q9PVG9
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POU-BOX PROTEIN BRAIN-2.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauiria; Aves; Neognathae; Galliformes; Phasianinae;
CC Coturnix.
NCBI_TaxID=93934;
[1]
SEQUENCE FROM N.A.
RP Liu Y., Xue J.X., Zhang W., Fu D.C., He R.Q., Xue Z.G.;
RA "qBrain-2" a Pou-box gene expressed in quail embryos.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RR -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: WITH OTHER HOMEOBOX PROTEINS.
DR EMBL; AF091043; AAA00040.1; -.
DR HSP; P14859; ILOC.
DR InterPro; IPRO01356; Homeobox.
DR InterPro; IPRO00327; POU.
DR Pfam; PF00046; homeobox_1.
DR PRINTS; PR00028; POU_DOMAIN.
DR ProDom; PD000383; POU; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00352; POU; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 431 AA; 43722 MW; IDC47E53PACC7D5 CRC64;

Query Match 35.7%; Score 80; DB 13; Length 431;
Best Local Similarity 48.8%; Pred. No. 0.87;
Matches 20; Conservative 0; Mismatches 7; Indels 14; Gaps 2;

Y 1 GGCGGEGPTLROWLAARAAGGGGGGGIEGPTLROWLAARA 41
||||| | : ||||
69 GGCGGGG-----GGGGGGGGGEAP----WAANA 95
RESULT 7
ID O24352 PRELIMINARY; PRT; 187 AA.
AC O24352
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MEN-4. PROTEIN.
GN MEN-4.
OS Silene latifolia.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
SEQUENCE FROM N.A.

R

TISSUE-WALE FLOWER;
RC MEDLINE=97377129; PubMed=9232878;
RX Scott C.P., Li Y., Robertson S.E., Willis M.E., Gilmarin P.M.;
RA "sex determination in Silene latifolia: y chromosome- and Ustilago
RT virulence-mediated effects during dioecious flower development.";
RL Plant Physiol. 114:969-979(1997).
DR EMBL; Y08776; CAAT0029.1; .
SQ SEQUENCE 187 AA; 18357 MW; FAFC6B5E7F7DFEE1 CRC64;

Query Match 35.5%; Score 79.5; DB 10; Length 187;
Best Local similarity 43.2%; Pred. No. 0.43;
Matches 19; Conservative 1; Mismatches 9; Indels 15; Gaps 2;

OY 1 GGCGGEGPTLROWLAARAAGGGGGGGI-----EGPTLRQW 36
||||| | : ||||
Db 84 GGCGKEG-----GGGGGGGGGANIPPRAAAGEGLIRSW 120
RESULT 8
ID Q9ASZ8 PRELIMINARY; PRT; 167 AA.
AC Q9ASZ8
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE P0416G11.17 PROTEIN.
GN P0416G11.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone-P0416G11.1";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002968; BAB39265.1; .
SQ SEQUENCE 167 AA; 17271 MW; 770B898A7294CC43 CRC64;

Query Match 35.3%; Score 79; DB 10; Length 167;
Best Local Similarity 40.5%; Pred. No. 0.43; 13; Indels 2; Gaps 1;
Matches 15; Conservative 7; Mismatches 7

OY 2 GGCGGEGPTLROWLAARAAGGGGGGGIEG--PTLRQW 36
||||| | : ||||
Db 66 GGCGVDNRKRQRRCQQDDGGGGGVGSFAFAAW 102
RESULT 9
ID O97344 PRELIMINARY; PRT; 650 AA.
AC O97344
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE INTERMEDIATE FILAMENT PROTEIN IF1.
OS Sagittaria elegans.
OC Eukaryota; Metazoa; Chaetognatha; Aphragmophora; Ctenodontina;
OC Sagittidae; Sagitta.
OX NCBI_TaxID=10231;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=99065768; PubMed=9847417;
RX Erber A., Riemer D., Bovenschulte M., Weber K.;
RA "Molecular Phylogeny of metazoan intermediate filament proteins."
RT J. Mol. Evol. 47:751-762(1998).
DL EMBL; AJ004932; CAB38182.1; .
DR InterPro; IPR001664; IF.
DR InterPro; IPR001322; IF_tail.
DR Pfam; PF000038; filament; 2.

```

AC Q19476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F15B9.5 PROPEIN.
GN F15B9.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 278013; CAB01420.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 500 AA; 33946 MW; 1416327086FE7CF6 CRC64;

Query Match 37.9%; Score 85; DB 5; Length 500;
Best Local Similarity 55.2%; Pred. No. 0.31;
Matches 16; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 2 GGGGIEGPTLRQWLAAAGGGGGGIEG 30
   ||| | | : | : | : | : | : | : | : |
Db 423 GGGAAAGMLGRFLNRRGGGGGGGGMG 451

RESULT 3
Q9LGC9
ID Q9LGC9 PRELIMINARY; PRT; 360 AA.
AC Q9LGC9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0462H08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002525; BAB07996.1; -.
DR InterPro; IPR000571; zf-CCCH.
DR Pfam; PF00642; zf-CCCH; 4.
DR SMART; SM00356; Znf_C3H1; 4.
SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 37.3%; Score 83.5; DB 10; Length 360;
Best Local Similarity 51.4%; Pred. No. 0.32;
Matches 18; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

QY 1 GGGG-----GIEGPTLRQWLAAAGGGGGGIEG 30
   || | | | : | | | | | : |
Db 16 GEGGASPDGTGLEPWRMLGSLGGGGGGGGGGDG 50

RESULT 4

```

```

Q23062
ID Q23062 PRELIMINARY; PRT; 374 AA.
AC Q23062;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 32.6 KDA PROTEIN.
GN T28H11.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Nelson J., Wohlmann P.;
RT "The sequence of C. elegans cosmid T28H11.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64609; AAB04604.2; -.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 32630 MW; CB0D4ECD7519D997 CRC64;

Query Match 37.1%; Score 83; DB 5; Length 374;
Best Local Similarity 59.3%; Pred. No. 0.38;
Matches 16; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 1 GGGGIEGPTLRQWLAAAGGGGGGGG 27
   ||||| | : : | |||||
Db 340 GGGGIPGQSV--YMGAGGGGGGGGG 364

RESULT 5
Q96755
ID Q96755 PRELIMINARY; PRT; 422 AA.
AC Q96755;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTERMEDIATE FILAMENT PROTEIN E1.
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7740;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99019308; PubMed=9804163;
RA Karabinos A., Riemer D., Erber A., Weber K.;
RT "Homologues of vertebrate type I, II and III intermediate filament
RT (IF) proteins in an invertebrate: the IF multigene family of the
RL cephalochordate Branchiostoma.";
RL FEBS Lett. 437:15-18(1998).
DR EMBL; AJ010294; CAA09068.1; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003489; Ribosomal_S30.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR01248; TYPE1KERATIN.

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run On: October 9, 2002, 08:52:16 ; Search time 14.6909 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-34
Perfect score: 224
Sequence: 1 GGGGGIEGPTLRQWLAAARAG.....GGGGGIEGPTLRQWLAAARA 41

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_19.*
- 1: sp-archaea.*
 - 2: sp-bacteria.*
 - 3: sp-fungi.*
 - 4: sp-human.*
 - 5: sp-invertebrate.*
 - 6: sp-mammal.*
 - 7: sp-mhc.*
 - 8: sp-organelle.*
 - 9: sp-phage.*
 - 10: sp-plant.*
 - 11: sp-rodent.*
 - 12: sp-virus.*
 - 13: sp-vertebrate.*
 - 14: sp-unclassified.*
 - 15: sp-rv.*
 - 16: sp-bacteriophage.*
 - 17: sp-archaeal.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.5	38.6	529	10 Q9ASE5	Q9ase5 oryza sativ
2	85	37.9	500	5 Q19476	Q19476 caenorhabdi
3	83.5	37.3	360	10 Q9LGC9	Q9lgc9 oryza sativ
4	83	37.1	374	5 Q23062	Q23062 caenorhabdi
5	81	36.2	422	5 Q96755	Q96755 brachyos
6	80	35.7	431	13 Q9PVG9	Q9pvg9 coturnix co
7	79.5	35.5	187	10 Q24352	Q24352 silene lati
8	79	35.3	167	10 Q9AS28	Q9as28 oryza sativ
9	79	35.3	650	5 Q97344	Q97344 sagitta ele
10	78	34.8	253	10 Q943K0	Q943k0 oryza sativ
11	78	34.8	308	10 Q9LGC9	Q9lgc9 oryza sativ
12	78	34.8	324	11 Q35295	Q35295 mus musculu
13	77	34.4	265	5 Q23347	Q23347 caenorhabdi
14	77	34.4	679	5 Q9W3E3	Q9w3e3 drosophila
15	76.5	34.2	194	5 Q96853	Q96853 schistosoma
16	76	33.9	377	13 Q9YHD0	Q9yhd0 petromyzon

17	76	33.9	1610	16 Q92KQ8	Q92kq8 rhizobium m
18	75.5	33.7	396	10 Q65450	Q65450 arabidopsis
19	75.5	33.7	642	13 Q9PUD8	Q9pud8 lampetra fl
20	75.5	33.7	687	13 Q9PUD9	Q9pud9 lampetra fl
21	75	33.5	250	5 Q24917	Q24917 echinometra
22	75	33.5	453	5 Q9N6M8	Q9nm8 drosophila
23	75	33.5	904	5 Q76271	Q76271 mytilus edu
24	75	33.5	1024	5 Q9VFM5	Q9vfm5 drosophila
25	74.5	33.3	113	10 Q942U6	Q942u6 oryza sativ
26	74.5	33.3	147	5 Q24348	Q24348 drosophila
27	74.5	33.3	308	5 Q95X69	Q95x69 caenorhabdi
28	74.5	33.3	344	5 Q9WLV3	Q9wlv3 drosophila
29	74.5	33.3	346	5 Q22037	Q22037 caenorhabdi
30	74.5	33.3	738	5 Q02402	Q02402 pinctada fu
31	74	33.0	228	10 Q9C7R3	Q9c7r3 arabidopsis
32	74	33.0	822	10 Q94HZ3	Q94hz3 oryza sativ
33	73.5	32.8	277	5 Q9W323	Q9w323 drosophila
34	73.5	32.8	535	10 Q942Q2	Q942q2 oryza sativ
35	73.5	32.8	827	5 Q960Z9	Q960z9 drosophila
36	73	32.6	236	5 Q24874	Q24874 echinometra
37	73	32.6	270	10 Q9SKZ1	Q9skz1 arabidopsis
38	73	32.6	296	10 Q9XHC4	Q9xhc4 arabidopsis
39	73	32.6	439	10 Q9SDK6	Q9sdk6 oryza sativ
40	73	32.6	453	5 Q9NGE6	Q9nge6 drosophila
41	73	32.6	454	5 Q9VBK9	Q9vbk9 drosophila
42	73	32.6	455	5 Q960C1	Q960c1 drosophila
43	73	32.6	499	12 Q9QMH3	Q9qmh3 parvo-like
44	72.5	32.4	237	5 Q9W2L8	Q9w2l8 drosophila
45	72	32.1	173	5 Q9VX65	Q9vx65 drosophila

ALIGNMENTS

RESULT 1

Q9ASE5 PRELIMINARY; PRT; 529 AA.

AC Q9ASE5; 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)

DE P0456F08.14 PROTEIN.

GN P0456F08.14.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaeae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

RT clone:P0456F08."

RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF002901; BAB39414.1;

DR InterPro; IPR002937; Amino_Oxidase.

DR InterPro; IPR002937; NAD_binding.

DR Pfam; PF01593; Amino_Oxidase; 1.

SQ SEQUENCE 529 AA; 55981 MW; 0A5DA55CDD076D24 CRC64;

Query Match 38.6%; Score 86.5; DB 10; Length 529;
Best Local Similarity 59.4%; Pred No. 0.23;
Matches 19; Conservative 3; Mismatches 7; Indels 3; Gaps 1;

QY 1 GGGGGIE--GPTLRQWLAAARAGGGGGGGGIE 29
|||||: ||: |||: |||||: |
Db 138 GGGGVEEYLRRLRAYQAARSAGGGGGGKE 169

RESULT 2
Q19476 PRELIMINARY; PRT; 500 AA.
ID Q19476


```

Query Match      30.6%; Score 68.5; DB 1; Length 759;
Best Local Similarity 40.5%; Pred. No. 9.2;
Matches 17; Conservative 5; Mismatches 13; Indels 7; Gaps 2;

QY      2 GGGGIEGP--TLRWLAARAGG-----GGGGIEGPTLRQW 36
      |||||::||::||| ||||| |||||
DB      231 GGGGLDAGGELKRLCFQDAGGGAGPPGGAGGAGGTWWRW 272

RESULT 29
TOP3_CAEEL      STANDARD;      PRT;      759 AA.
ID   TOP3_CAEEL
AC   O61660;
DC   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DD   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA topoisomerase III (EC 5.99.1.2).
GN   Top3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
RN   [1]
RS   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Kim Y.-C., Koo H.-S.;
RT   "cDNA cloning and overexpression of Caenorhabditis elegans DNA
  topoisomerase III.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC   -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
  DNA, followed by passage and rejoining.
CC   -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
  FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@lsb-sib.ch).
CC   -----
DR   EMBL; AF057032; AAC13567.1;
DR   InterPro; IPR003601; DNATopI_ATP_bind.
DR   InterPro; IPR003602; DNATopI_DNA_bind.
DR   InterPro; IPR000380; Pro_topoisomerase.
DR   InterPro; IPR002936; Toprim.
DR   Pfam; PF01131; Topoisom_bac; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   PRINTS; PR00417; PRTPISMRASEI.
DR   SMART; SM00437; TOPIAC; 1.
DR   SMART; SM00436; TOPIbc; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   SMART; SM00396; TOPOISOMERASE_I_PROK; 1.
DR   PROSITE; PS00396; TOPOISOMERASE; DNA-binding.
KW   Isomerase; Topoisomerase; DNA-binding.
FT   ACT_SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).
FT   SEQUENCE 759 AA; 85438 MW; 3D862412D72946BD CRC64;
SQ
Query Match      30.6%; Score 68.5; DB 1; Length 759;
Best Local Similarity 51.6%; Pred. No. 9.2;
Matches 16; Conservative 0; Mismatches 8; Indels 7; Gaps 1;

QY      1 GGGGIEGPTLRWLAARAGGGGGGGIEGP 31
      |||||::||| | ||||| |||||
DB      619 GPGGGGGP-----PRGPGGGGGGGGTGP 642

RESULT 30
CIKF_MOUSE      STANDARD;      PRT;      769 AA.
ID   CIKF_MOUSE
AC   Q63959; Q62088;
DC   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DD

```


KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
KW Glycoprotein; Repeat.

FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 49 LACCASE.
FT CHAIN 50 606
FT PROPEP 607 619
FT DOMAIN 84 207
FT DOMAIN 216 373 PLASTOCYANIN-LIKE 1.
FT METAL 431 566 PLASTOCYANIN-LIKE 2.
FT METAL 144 144 PLASTOCYANIN-LIKE 3.
FT METAL 146 146 COPPER (TYPE 2) (PROBABLE).
FT METAL 189 189 COPPER (TYPE 3) (PROBABLE).
FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
FT METAL 477 477 COPPER (TYPE 1) (PROBABLE).
FT METAL 480 480 COPPER (TYPE 2) (PROBABLE).
FT METAL 482 482 COPPER (TYPE 3) (PROBABLE).
FT METAL 548 548 COPPER (TYPE 3) (PROBABLE).
FT METAL 549 549 COPPER (TYPE 1) (PROBABLE).
FT METAL 550 550 COPPER (TYPE 3) (PROBABLE).
FT METAL 554 554 COPPER (TYPE 1) (PROBABLE).
FT METAL 559 559 COPPER (TYPE 1) (PROBABLE).
FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 619 AA; 68198 MW; FDED6D78B65048E3 CRC64;

Query Match 30.6%; Score 68.5; DB 1; Length 619;
Best Local Similarity 48.7%; Pred. No. 7.7; Indels 7; Gaps 2;
Matches 19; Conservative 0; Mismatches 13;

OY 5 GIEGPTL-----ROWLAARAGGGGGGIEGPTLRQ-W 36
Db 26 GTEGVNLLTPVDKRDQSOAERYGGGGGNCSPNRCQW 64

RESULT 27
LAC2_NEUCR

ID LAC2_NEUCR STANDARD; PRT; 619 AA.
AC P10574;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
DE (urishiol oxidase) (Laccase allele TS).
GN LACC.

OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]

RP SEQUENCE FROM N.A.
RA MEDLINE=88087214; PubMed=2961749;

RX Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
RT "Characterization of two allelic forms of Neurospora crassa laccase.
RT Amino- and carboxyl-terminal processing of a precursor.";
RL J. Biol. Chem. 263:885-896(1988).

CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.

CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.

CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M18334; AAA33592.1; -
DR PIR; B28523; KSNCLT.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu_oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
KW Glycoprotein; Repeat.

FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 49 LACCASE.
FT CHAIN 50 606
FT PROPEP 607 619
FT DOMAIN 84 207
FT DOMAIN 216 373 PLASTOCYANIN-LIKE 1.
FT METAL 431 566 PLASTOCYANIN-LIKE 2.
FT METAL 144 144 PLASTOCYANIN-LIKE 3.
FT METAL 146 146 COPPER (TYPE 2) (PROBABLE).
FT METAL 189 189 COPPER (TYPE 3) (PROBABLE).
FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
FT METAL 477 477 COPPER (TYPE 1) (PROBABLE).
FT METAL 480 480 COPPER (TYPE 2) (PROBABLE).
FT METAL 482 482 COPPER (TYPE 3) (PROBABLE).
FT METAL 548 548 COPPER (TYPE 3) (PROBABLE).
FT METAL 549 549 COPPER (TYPE 1) (PROBABLE).
FT METAL 550 550 COPPER (TYPE 3) (PROBABLE).
FT METAL 554 554 COPPER (TYPE 1) (PROBABLE).
FT METAL 559 559 COPPER (TYPE 1) (PROBABLE).
FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 619 AA; 68120 MW; 0BB6CCDE18841145 CRC64;

Query Match 30.6%; Score 68.5; DB 1; Length 619;
Best Local Similarity 48.7%; Pred. No. 7.7; Indels 7; Gaps 2;
Matches 19; Conservative 0; Mismatches 13;

OY 5 GIEGPTL-----ROWLAARAGGGGGGIEGPTLRQ-W 36
Db 26 GTEGVNLLTPVDKRDQSOAERYGGGGGNCSPNRCQW 64

RESULT 28

CIKF_HUMAN
ID CIKF_HUMAN STANDARD; PRT; 757 AA.
AC Q14003;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-gated potassium channel protein Kv3.3 (KSHIID).
GN KCNC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Lens epithelium;
RX MEDLINE=20179629; PubMed=10712820;
RA Rae J.L., Shepard A.R.;
RT "Kv3.3 potassium channels in lens epithelium and corneal
RT endothelium.";
RL Exp. Eye Res. 70:339-348(2000).
RN [2]
RP SEQUENCE OF 291-651 FROM N.A.

us-09-422-838c-34.rsp

Wed Oct 9 10:30:20 2002

```

FT DOMAIN 312 333 LEUCINE-ZIPPER.
SQ SEQUENCE 369 AA; 38457 MW; 288B464708DA6C7D CRC64;

Query Match 30.6%; Score 68.5; DB 1; Length 369;
Best Local Similarity 53.3%; Pred. No. 5;
Matches 16; Conservative 1; Mismatches 6; Indels 7; Gaps 2;

OY 1 GGGGGIEGPTLRQWLAARAGGGGGGGGIEG 30
DB 215 GAGGG--GPA-----SAGGGGGGGGGGTAG 237

RESULT 25
WASP_MOUSE STANDARD; PRT; 520 AA.
ID WASP_MOUSE
AC P70315;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Wiskott-Aldrich syndrome protein homolog (WASP).
GN WAS OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=96115600; PubMed=8666397;
RA Lerray J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Denahy V., Godfrey V.L., Wilkinson J.E., Francke U.;
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sf) mutation on
RT the X chromosome." (1995).
RL Genomics 29:471-477(1995).
CC -!- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION
CC (BY SIMILARITY).
CC -!- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -!- SIMILARITY: CONTAINS 1 GBD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC EMBL; U54788; AAC52556.1; -.
CC MGD; MGI:105059; Was.
CC InterPro: IPR000095; PAK_box_P21_Rho_binding.
CC InterPro: IPR000697; RanBP1_WASP.
CC InterPro: IPR001960; WH1.
CC InterPro: IPR003124; WH2.
CC Pfam; PF00786; PBD; 1.
CC Pfam; PF00568; WH1; 1.
CC Pfam; PF02205; WH2; 1.
CC Pfam; SM00285; PBD; 1.
CC SMART; SM00461; WH1; 1.
CC SMART; SM00246; WH2; 1.
CC PROSITE; PS50108; GBD; 1.
KW Repeat.
FT DOMAIN 41 147 WH1.
FT DOMAIN 240 259 GBD.
FT REPEAT 354 363 GRSGPLPPXP MOTIF 1.
FT REPEAT 393 402 GRSGPLPPXP MOTIF 2.
FT DOMAIN 162 167 POLY-PRO.
FT DOMAIN 314 321 POLY-PRO.
FT DOMAIN 324 341 POLY-GLY.
FT DOMAIN 368 373 POLY-PRO.

```

```

FT DOMAIN 376 390 POLY-PRO.
FT DOMAIN 384 390 POLY-PRO.
FT DOMAIN 397 403 POLY-PRO.
FT DOMAIN 408 424 POLY-PRO.
FT DOMAIN 503 520 ASP/GLD-RICH (ACIDIC).
SQ SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Query Match 30.6%; Score 68.5; DB 1; Length 520;
Best Local Similarity 47.1%; Pred. No. 6.7;
Matches 16; Conservative 0; Mismatches 5; Indels 13; Gaps 1;

OY 1 GGGGGIEGPTLRQWLAARAGGGGGGGGIEGPTLR 34
DB 325 GGGGG-----GGGGGGGGGGGQPLR 345

RESULT 26
LACL_NEUCR STANDARD; PRT; 619 AA.
ID LACL_NEUCR
AC P06811;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase precursor (EC 1.10.3.2) (benzenediol:oxygen oxidoreductase)
DE (Urishiol oxidase) (Laccase allele OR).
GN LACC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=88087214; PubMed=2961749;
RA Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
RA "Characterization of two allelic forms of Neurospora crassa laccase.
RT Amino- and carboxyl-terminal processing of a precursor.";
RL J. Biol. Chem. 263:885-896(1988).
RN [2]
RN SEQUENCE OF 379-619 FROM N.A.
RP MEDLINE=87067412; PubMed=2947240;
RA Germann U.A., Lerch K.;
RA "Isolation and partial nucleotide sequence of the laccase gene from
RT Neurospora crassa: amino acid sequence homology of the protein to
RT human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).
CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.
CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC EMBL; M14554; AAA33590.1; -.
CC PIR; A28523; KSNCLQ.
CC PIR; A29762; A29762.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR002355; MultiCu_oxidase.
CC Pfam; PF00394; Cu-oxidase; 3.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
CC PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.

```

CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Plank J.L., Reineke J.C., Willson T.M., Hsieh T.-S.;
 RT "Drosophila melanogaster topoisomerase III alpha";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam K.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: WEAKLY RELAXES NEGATIVE SUPERCOILS AND DISPLAYS A
 CC -!- DISTINCT PREFERENCE FOR BINDING SINGLE-STRANDED DNA.
 CC -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
 CC DNA, followed by passage and rejoining.
 CC -!- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
 CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF255733; AAF71288.1;
 CC EMBL; AE003663; AAF53813.1; ALT_SEQ.
 CC FlyBase; FBgn0040268; Top3-alpha.
 DR InterPro; IPR003601; DNATopI_Atp_bind.

DR InterPro; IPR003602; DNATopI_DNA_bind.
 DR InterPro; IPR000380; Pro_topoisomerase.
 DR InterPro; IPR002936; Toprim.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF01131; Topoisom_bac; 1.
 DR Pfam; PF01751; Toprim; 1.
 DR Pfam; PF01396; zfc4_Topoism; 1.
 DR PRINTS; PR00417; PRTPISMRASEI.
 DR SMART; SM00437; TOPIAC; 1.
 DR SMART; SM00436; TOP1Bc; 1.
 DR SMART; SM00493; TOPRIM; 1.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
 KW Isomerase; Topoisomerase; DNA-binding.
 KW ACT_SITE 336 356 DNA_CLEAVAGE (BY SIMILARITY).
 FT SEQUENCE 1250 AA; 136137 MW; 28809F70B3DB75E CRC64;
 SQ
 Query Match 30.8%; Score 69; DB 1; Length 1250;
 Best Local Similarity 46.7%; Pred No. 13;
 Matches 14; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
 QY 1 GGGGGTEGPTLRQWLAARAGGGGGGGG 30
 Db 780 GGGGGGAGGAGGSGGSGGGGGGGG 809
 ID MAF2_RAT STANDARD; PRT; 369 AA.
 AC P54844;
 DT 01-OCT-1996 (Rel. 34; Created)
 DT 01-OCT-1996 (Rel. 34; Last sequence update)
 DT 30-MAY-2000 (Rel. 39; Last annotation update)
 DE Transcription factor MAF2 (Proto-oncogene C-MAF).
 GN MAF2 OR MAF.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WISTAR;
 RA MEDLINE=97190228; PubMed=9038383;
 RA Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,
 RA Kuboki Y., Nishizawa M., Nishi S.;
 RT "Rat maf related genes: specific expression in chondrocytes, lens and
 RT spinal cord";
 RL Oncogene 14:745-750(1997).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U56242; AAB50063.1;
 CC HSSP; P05412; IJUN.
 DR InterPro; IPR001871; bZIP.
 DR Pfam; PF03131; bZIP_Maf; 1.
 DR SMART; SM00338; BRUZ; 1.
 KW Proto-oncogene; Transcription regulation; DNA-binding; Activator;
 KW Nuclear protein.
 FT DOMAIN 139 146 POLY-ALA.
 FT DOMAIN 169 173 POLY-ALA.
 FT DOMAIN 180 187 POLY-HIS.
 FT DOMAIN 191 194 POLY-HIS.
 FT DOMAIN 212 220 POLY-GLY.
 FT DOMAIN 225 234 POLY-GLY.
 FT DOMAIN 284 310 BASIC MOTIF.

```
DR SMART: SM00338; BRLZ; 1.
KW Oncogene; Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 163 167 POLY-ALA.
FT DOMAIN 173 182 HIS-RICH.
FT DOMAIN 184 191 POLY-RICH.
FT DOMAIN 214 223 POLY-GLY.
FT DOMAIN 233 240 POLY-GLY.
FT DNA_BIND 274 300 BASIC MOTIF.
FT DOMAIN 302 323 LEUCINE-ZIPPER.
SQ SEQUENCE 369 AA; 38992 MW; F386B220ACES0FF6 CRC64;

Query Match 31.0%; Score 69.5; DB 1; Length 369;
Best Local Similarity 44.7%; Pred. No. 4;
Matches 17; Conservative 2; Mismatches 8; Indels 11; Gaps 1;

QY 1 GGGGGGEG-----PTLQWLAAARAGGGGGG 27
      ||||| | | | | | | | | | | | | | | | | |
Db 184 GGGGGGGHGHGAPGAPPSSASSAAGSGGGGGG 221

RESULT 21
ROAL_SCHAM STANDARD; PRT; 342 AA.
AC P21522;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein A1, A2/B1 homolog.
GN HNRNP.
OS Schistocerca americana (American grasshopper).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridoidea; Acridoidea; Acrididae; Schistocerca.
OX NCBI_TaxID=7009;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195067; PubMed=1901646;
RA Ball E.E., Rehm E.J., Goodman C.S.;
RT "Cloning of a grasshopper cDNA coding for a protein homologous to the
RL Nucleic Acids Res. 19:397-397(1991).
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF RIBONUCLEOSOMES.
CC -!- DEVELOPMENTAL STAGE: ABUNDANT IN EMBRYOS YOUNGER THAN 50% OF
CC EMBRYONIC DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE A/B GROUP OF HNRNP, WHICH ARE BASIC AND
CC GLY-RICH PROTEINS.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54670; CAA38481.1; -
CC DR PIR; S13513; S13513.
CC DR PIR; S14432; S14432.
CC DR HSSP; P09651; IUP1.
CC DR InterPro; IPR000504; RRM.
CC DR Pfam; PF00076; rrm; 2.
CC DR SMART; SM00360; RRM; 2.
CC DR PROSITE; PS00102; RRM; 2.
CC DR Nucleic acid protein; RNA-binding; Ribonucleoprotein; Repeat.
KW DOMAIN 17 93 RNA-BINDING (RRM) 1.
FT DOMAIN 108 185 RNA-BINDING (RRM) 2.
FT DOMAIN 187 339 GLY-RICH.
SQ SEQUENCE 342 AA; 35404 MW; F8426A21C8125DD8 CRC64;

Query Match 30.8%; Score 69; DB 1; Length 342;
Best Local Similarity 50.0%; Pred. No. 4.2;

Matches 17; Conservative 1; Mismatches 10; Indels 6; Gaps 2;

QY 1 GGGGGEGPTLQWLAAARAGG-----GGGGGGIEG 30
      ||||| | | | | | | | | | | | | | | | | |
Db 223 GGGGGGGAD--PWENGRGGGGDRWGGGGGGMG 254

RESULT 22
GRPL_PETHY STANDARD; PRT; 384 AA.
AC P09789;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Glycine-rich cell wall structural protein 1 precursor.
GN GRP-1.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RA Condit C.M., Meagher R.B.;
RT "A gene encoding a novel glycine-rich structural protein of petunia.";
RL Nature 323:178-181(1986).
CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -!- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.
CC -!- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF
CC FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
CC -!- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
CC FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40
CC AA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X04335; CAA27866.1; -
CC DR PIR; A26099; A26099.
CC DR HSSP; P30129; 4DPV.
CC KW Cell wall; Structural protein; Repeat; Signal.
CC FT SIGNAL 1 27
CC FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL
CC FT DOMAIN 41 384 GLY-RICH.
CC SQ SEQUENCE 384 AA; 28777 MW; CB541C549417D18C CRC64;

Query Match 30.8%; Score 69; DB 1; Length 384;
Best Local Similarity 46.7%; Pred. No. 4.6;
Matches 14; Conservative 2; Mismatches 4; Indels 10; Gaps 1;

QY 1 GGGGGEGPTLQWLAAARAGGGGGGGIEG 30
      ||||| | | | | | | | | | | | | | | | | |
Db 84 GGGGGGAG-----GGGAGGGGGG 103

RESULT 23
TP3A_DROME STANDARD; PRT; 1250 AA.
AC Q9NG98; Q9VIV1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase III alpha (EC 5.99.1.2).
OS TOP3-ALPHA OR CGI1023.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
```

```

DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 2.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
FT DNA-BIND 140 199 HOMEBOX.
FT DOMAIN 313 326 OAR.
FT DOMAIN 60 86 POLY-GLY.
FT VARSPIC 235 246 MISSING (IN ISOFORM SHOX2B).
FT CONFLICT 115 115 E -> EGRKPKAEVQATLLPGFAFRL (IN REF. 2).
FT CONFLICT 125 125 E -> D (IN REF. 1; CAA05341).
FT CONFLICT 244 244 P -> S (IN REF. 2).
FT CONFLICT 312 312 D -> N (IN REF. 2 AND 3).
FT CONFLICT 325 325 H -> L (IN REF. 3; AAC39663).
SQ SEQUENCE 331 AA; 34964 MW; 55431B073B3B2250 CRC64;

Query Match 31.2%; Score 70; DB 1; Length 331;
Best Local Similarity 50.0%; Pred. No. 3.3;
Matches 15; Conservative 1; Mismatches 4; Indels 10; Gaps 1;

Qy 1 GGGGIEGPTLROWLAARAGGGGGGIEG 30
Db 61 GGGGGGGG-----GGGGGGGGG 80

RESULT 19
R887_DROME STANDARD; PRT; 386 AA.
AC P48810;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein 87F (HRP36.1 protein) (P11
DE protein).
GN HRB87F OR HRP36.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phlebotomidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R, AND CANTON-S; TISSUE=Ovary;
RX MEDLINE=91187645; PubMed=1849257;
RA Haynes S.R., Johnson D., Raychaudhuri G., Beyer A.L.;
RT "The Drosophila Hrb87F gene encodes a new member of the A and B hnRNP
RT protein group.";
RL Nucleic Acids Res. 19:25-31(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=92112968; PubMed=1730754;
RA Matunis E.L., Matunis M.J., Dreyfuss G.;
RT "Characterization of the major hnRNP proteins from Drosophila
RT melanogaster.";
RL J. Cell Biol. 116:257-269(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=Embryo;
RX MEDLINE=92020124; PubMed=1717937;
RA Hovemann B.T., Dessen E., Mechler H., Mack E.;
RT "Drosophila snRNP associated protein P11 which specifically binds to
RT heat shock snRNP 93D reveals strong homology with hnRNP core protein
RT Al.";
RL Nucleic Acids Res. 19:4909-4914(1991).
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF RIBONUCLEOSOMES. COULD BE
CC NEEDED TO ORGANIZE A CONCENTRATION GRADIENT OF A DORSALIZING
CC MORPHOGEN (Dm) ORIGINATING IN THE GERMINAL VESICLE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND/OR CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; X54803; CAA38574.1; -.
EMBL; X62636; CAA44502.1; -.
EMBL; X59691; CAA42212.1; -.
DR HSSP; P09651; IUP1.
DR FlyBase; FBgn0004237; Hrb87F.
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 2.
KW RNA-binding; Nuclear protein; Ribonucleoprotein; Repeat;
KW Alternative splicing.
FT DOMAIN 24 101 RNA-BINDING (RRM) 1.
FT DOMAIN 115 192 RNA-BINDING (RRM) 2.
FT VARSPIC 315 374 MISSING (IN ISOFORM HRP36.1).
FT CONFLICT 271 271 S -> T (IN REF. 3).
SQ SEQUENCE 386 AA; 39557 MW; 2036C04D01E3AFD7 CRC64;

Query Match 31.2%; Score 70; DB 1; Length 386;
Best Local Similarity 51.9%; Pred. No. 3.8;
Matches 14; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 GGGGIEGPTLROWLAARAGGGGGGGG 27
Db 290 GGGGGYGGGNGSWGNGGGGGGGGG 316

RESULT 20
TMAF_AVIS4 STANDARD; PRT; 369 AA.
ID P23091;
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transforming protein Maf.
GN V-MAF.
OS Avian musculoaponeurotic fibrosarcoma virus AS42.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11873;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046665; PubMed=2554284;
RA Nishizawa M., Kataoka K., Goto N., Fujiwara K.T., Kawai S.;
RT "v-maf, a viral oncogene that encodes a 'leucine zipper' motif.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7711-7715(1989).
CC -!- FUNCTION: MIGHT BE A TRANSCRIPTIONAL TRANS-ACTIVATOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: INDUCES MUSCULOAPONEUROTIC FIBROSARCOMA IN CHICKENS.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A ENV-MAF
CC POLYPROTEIN.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; M26769; AAA42377.1; -.
DR PIR; B33975; TPEVAF.
DR HSSP; P05412; LTUN.
DR TRANSFAC; T01430; -.
DR InterPro; IPR001871; bZIP.
DR Pfam; PF03131; bZIP_Maf; 1.

```

RA Kidd K.K., Busyгина V., Demille M.M.C., Speed W.C., Ruggeri V.,
 RA Kidd J.R., Pakstis A.J.;
 RT "Overall linkage disequilibrium in 33 populations for highly
 RT informative multiseite haplotypes spanning the HOXB gene cluster.";
 RL Am. J. Hum. Genet. 67:235-235(2000).
 RN
 RN SEQUENCE OF 188-253 FROM N.A.
 RC TISSUE=Placenta;
 RX GIampolo A., Acampora D., Zappavigna V., Pannese M.,
 RA D'Esposito M., Care A., Faiella A., Stornaiuolo A., Russo G.,
 RA Simeone A., Boncinelli E., Peschle C.;
 RT "Differential expression of human HOX-2 genes along the anterior-
 RT posterior axis in embryonic central nervous system";
 RT Differentiation 40:191-197(1989).
 RN
 RN SEQUENCE OF 188-253 FROM N.A.
 RX MEDLINE=90215256; PubMed=2576652;
 RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,
 RA Gaudino G., Stornaiuolo A., Cafiero M., Faiella A., Simeone A.;
 RT "Organization of human class I homeobox genes";
 RL Genome 31:745-756(1989).
 CC
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT
 CC 5-9 WEEKS FROM CONCEPTION.
 CC
 CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X16667; CAA34657.1; -
 CC EMBL; U59298; AAD10852.1; -
 CC EMBL; AF287967; AAG31555.1; -
 CC EMBL; X16175; CAA34297.1; -
 CC PIR; S07543; WJHU26.
 CC PIR; D37042; D37042.
 CC HSSP; P02833; LSAN.
 CC TRANSFAC; T01723; -
 CC MIM; 142966; -
 CC InterPro; IPR001827; Antennapedia.
 CC InterPro; IPR001356; Homeobox.
 CC Pfam; PF00046; homeobox; 1.
 CC PRINTS; PR00025; ANTENNAPEDIA.
 CC PRINTS; PR00024; HOMEBOX.
 CC SMART; SM00389; HOX; 1.
 CC PROSITE; PS00027; HOMEBOX_1; 1.
 CC PROSITE; PS00032; ANTENNAPEDIA; 1.
 CC PROSITE; PS0071; HOMEBOX_2; 1.
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 129 134 ANTP-TYPE HEXAPEPTIDE.
 FT DOMAIN 154 178 GLY-RICH.
 FT DNA_BIND 188 247 HOMEBOX.
 FT CONFLICT 199 200 QL -> HV (IN REF. 2).
 FT SEQUENCE 431 AA; 44344 MW; 941706EDCC2975E5 CRC64;
 Query Match 31.5%; Score 70.5; DB 1; Length 431;
 Best Local Similarity 48.4%; Pred. No. 3.7;
 Matches 15; Conservative 2; Mismatches 7; Gaps 1;
 QY 1 GGGGGGEGPTLROWLAARAGGGGGGGGIE3P 31
 Db 159 GGGGGGSG-----GSGGGGGGGGGGDKSP 182

RESULT 18
 SHX2_HUMAN
 ID SHX2_HUMAN STANDARD; PRT; 331 AA.
 AC O60902; O60903; O60465; O60467;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Short stature homeobox protein 2 (Paired-related homeobox protein
 DE SHOT) (Homeobox protein Ogl2X).
 OS Homo sapiens (Human).
 GN SHOX2 OR SHOT OR OGL2X.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Fibroblast;
 RX MEDLINE=98151525; PubMed=9482898;
 RA Blaschke R.J., Monaghan A.P., Schiller S., Schechinger B., Rao E.,
 RA Padilla-Nash H., Ried T., Rappold G.A.;
 RT "SHOT, a SHOX-related homeobox gene, is implicated in craniofacial,
 RT brain, heart, and limb development.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:2406-2411(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 116-331 FROM N.A. (ISOFORM 2).
 RC TISSUE=Craniofacial;
 RX MEDLINE=98133920; PubMed=9466998;
 RA Semina E.V., Reiter R.S., Murray J.C.;
 RT "A new human homeobox gene OGL2X is a member of the most conserved
 RT homeobox gene family and is expressed during heart development in
 RT mouse.";
 RL Hum. Mol. Genet. 7:415-422(1998).
 CC
 CC -1- FUNCTION: May be a growth regulator and have a role in specifying
 CC neural systems involved in processing somatosensory information,
 CC as well as in face and body structure formation.
 CC
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/SHOX2A/SHOTA (shown here) and
 CC 2/SHOX2B/SHOTB/OGL2XB; are produced by alternative splicing.
 CC
 CC -1- TISSUE SPECIFICITY: Expressed in heart, skeletal muscle, liver,
 CC lung, bone marrow fibroblast, pancreas and placenta.
 CC
 CC -1- DEVELOPMENTAL STAGE: Expressed during craniofacial development as
 CC well as in heart.
 CC
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
 CC "BICOID" SUBFAMILY.
 CC
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ002367; CAA05341.1; ALT_INIT.
 CC EMBL; AJ002368; CAA05342.1; ALT_INIT.
 CC EMBL; BC008829; AAH08829.1; -
 CC EMBL; AF022654; AAC39662.1; ALT_INIT.
 CC EMBL; AF023203; AAC39663.1; -
 CC HSSP; P06601; 1FUJ.
 CC MIM; 602504; -
 CC InterPro; IPR000047; HTH_repressr.
 CC InterPro; IPR001356; Homeobox.
 CC InterPro; IPR003654; OAR_domain.
 CC Pfam; PF00046; homeobox; 1.
 CC PRINTS; PR00024; HOMEBOX.
 CC PRINTS; PR00031; HTHREPRESSR.
 CC SMART; SM00389; HOX; 1.

```

RP SEQUENCE FROM N.A.
RA MEDLINE=96144300; PubMed=8566805;
RX Thompson S.A., Gollightly E.J., Yaver D.S.;
RT "Nucleotide sequence of the Aspergillus niger srpA gene.";
RL Gene 167:337-338(1995).
CC -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECRETORY PROTEIN
CC WHEN THEY EMERGE FROM THE RIBOSOMES AND TRANSFERS THEM TO TRAM
CC (TRANSLATING CHAIN-ASSOCIATING MEMBRANE PROTEIN).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DOMAIN: HAS A TWO DOMAIN STRUCTURE. THE G-DOMAIN BINDS GTP; THE
CC M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE
CC SIGNAL SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L38317; AAB04946.1;
DR HSSP: 007347; 2FFH.
DR InterPro: IPR000897; SRP54.
DR InterPro: IPR004125; SRP_SPB.
DR Pfam: PF00448; SRP54; 1.
DR Pfam: PF02881; SRP54; 1.
DR Pfam: PF02978; SRP_SPB; 1.
DR ProDom: PD000819; SRP54; 1.
DR ProSITE: PS00300; SRP54; 1.
KW Signal recognition particle;
FT DOMAIN 1 296
FT NP_BIND 109 116 M-DOMAIN.
FT NP_BIND 191 195 GTP (BY SIMILARITY).
FT NP_BIND 249 252 GTP (BY SIMILARITY).
FT DOMAIN 468 474 POLY-GLY.
FT DOMAIN 493 501 POLY-GLY.
FT DOMAIN 522 533 POLY-GLY.
SQ SEQUENCE 534 AA; 57117 MW; 1B21E3A48CCB3BA6 CRC64;

Query Match 31.7%; Score 71; DB 1; Length 534;
Best Local Similarity 42.4%; Pred. No. 4;
Matches 14; Conservative 5; Mismatches 8; Indels 6; Gaps 1;

QY 1 GGGGIGEG-----PTLRQLAARAGGGGGGGG 27
Db 497 GGGGGLPLGLMDLQSMMSQMSGLMGGGGGGG 529

RESULT 16
SUS_DROME
ID SUS_DROME STANDARD; PRT; 1322 AA.
AC P22293;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Suppressor of sable protein.
GN SU(S).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=91117256; PubMed=1703632;
RA Voelker R.A., Gibson W., Graves J.P., Sterling J.F., Eisenberg M.T.;
RT "The Drosophila suppressor of sable gene encodes a polypeptide with
RT regions similar to those of RNA-binding proteins.";
RL Mol. Cell. Biol. 11:894-905(1991).
```

```

[2]
RN SEQUENCE OF 1-9 FROM N.A.
RA MEDLINE=91169252; PubMed=1963868;
RX Voelker R.A., Graves J.P., Gibson W., Eisenberg M.T.;
RT "Mobile element insertions causing mutations in the Drosophila
RT suppressor of sable locus occur in Dnase I hypersensitive subregions
RT of 5'-transcribed nontranslated sequences.";
RL Genetics 126:1071-1082(1990).
CC -!- FUNCTION: AFFECTS THE TRANSCRIPT LEVELS OF THOSE ALLELES THAT IT
CC SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: AT ALL STAGES.
CC -!- SIMILARITY: HAS REGIONS SIMILAR TO THOSE OF RNA-BINDING PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M57889; AAA28920.1;
DR EMBL: X59364; CAA42010.1;
DR PIR: A39612; A39612.
DR FlyBase: FBgn0003575; su(s).
DR InterPro: IPR000571; 2F-CCCH.
DR Pfam: PF00642; zf-CCCH; 2.
KW RNA-binding; Nuclear protein.
FT DOMAIN 138 327 HIGHLY CHARGED DOMAIN.
FT DOMAIN 446 474 GLN-RICH (OPA-REPEAT).
FT DOMAIN 1087 1162 RNA-BINDING (BY SIMILARITY).
SQ SEQUENCE 1322 AA; 143555 MW; D5F534EB5702EA08 CRC64;

Query Match 31.7%; Score 71; DB 1; Length 1322;
Best Local Similarity 47.1%; Pred. No. 8.7;
Matches 16; Conservative 1; Mismatches 9; Indels 8; Gaps 1;

QY 2 GGGGIEGPTLRQLAARAGGGGGGGGIEGPTLRQ 35
Db 1149 GGGGDSG-----GGVGGGGGGGVLPNLSQ 1174

RESULT 17
HXB3_HUMAN
ID HXB3_HUMAN STANDARD; PRT; 431 AA.
AC P14651; P17484; O95615;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-B3 (Hox-2G) (Hox-2.7).
GN HOXB3 OR HOX2G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90098876; PubMed=2574852;
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family.";
RL Nucleic Acids Res. 17:10385-10402(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Sauvageau G., Thorsteinsdottir U., Hough M.R., Hugo P., Lawrence H.J.,
RA Largman C., Humphries R.K.;
RT "Deregulated expression of HoxB3 in hematopoietic cells causes
RT defective development of alpha beta T lymphocytes and progressive
RT myeloproliferation.";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

Wed Oct 9 10:30:20 2002

RL Mol. Cell. Biol. 15:4562-4571(1995).
 RN [2]
 RC SEQUENCE FROM N.A. (ISOFORM 1).
 RP STRAIN=CANTON-S;
 RX MEDLINE=95223793; PubMed=7708500;
 RA Stoolow D.T., Haynes S.R.;
 RT "Cabeza", a *Drosophila* gene encoding a novel RNA binding protein,
 RT shares homology with EWS and TLS, two genes involved in human sarcoma
 RT formation.";
 RL Nucleic Acids Res. 23:835-843(1995).
 RN [3]
 RC SEQUENCE FROM N.A. (ISOFORM 1).
 RP STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Manatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA Amanatides P.G., Scher S.E., Richards S., Ashburner M., Henderson S.N.,
 RA George R.A., Lewis S.E., Riddle S.E., Zhang Q., Chen L.X.,
 RA Sutton R.C., Rogers J., H.C., Blazek R.G., Champ M., Pfeiffer B.D.,
 RA Brandon R.C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Abell J.F., Agbayani A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liao X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stاپleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Wu D.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE OF 39-404 FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=87175568; PubMed=3031652;
 RA Haynes S.R., Rebert M.L., Mozer B.A., Forquignon F., David I.B.;
 RT "pen repeat sequences are GGN clusters and encode a glycine-rich
 RT domain in a *Drosophila* cDNA homologous to the rat helix destabilizing
 RT protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:1819-1823(1987).
 CC -1- FUNCTION: MAY PARTICIPATE IN A FUNCTION COMMON TO THE EXPRESSION
 CC OF MOST GENES TRANSCRIBED BY RNA POLYMERASE II.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: 1 AND 2 (SHOWN HERE);
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. ENRICHED IN THE BRAIN AND CENTRAL
 CC NERVOUS SYSTEM DURING EMBRYOGENESIS. ENRICHED IN THE ADULT HEAD.

CC EMBRYOS CONTAIN BOTH TYPE 1 AND TYPE 2 ISOFORMS, WHEREAS LATER IN
 CC DEVELOPMENT (HEADS AND TORSOS) ONLY THE TYPE 2 ISOFORM IS
 CC DETECTED.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING EMBRYO FROM THE
 CC EARLIEST STAGES OF CELLULARIZATION AND IS SUBSEQUENTLY FOUND IN
 CC MANY CELL TYPES.
 CC -1- MISCELLANEOUS: 'CABEZA' MEANS 'HEAD' IN SPANISH.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
 CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U13178; AAA86955.1; -
 CC EMBL; L37083; AAC41563.1; -
 CC EMBL; AE003501; AAC48578.1; -
 CC EMBL; M15765; AAA70425.1; -
 CC FlyBase; FBgn0011571; caz.
 CC InterPro; IPR000504; RRM.
 CC InterPro; IPR001876; Znf-RanBP.
 CC Pfam; PF00076; rrm; 1.
 CC Pfam; PF00641; zf-RanBP; 1.
 CC SMART; SM00360; RRM; 1.
 CC SMART; SM00547; Znf-RBZ; 1.
 CC PROSITE; PS0102; RRM; 1.
 CC PROSITE; PS01358; ZF-RANBP2_1; 1.
 CC PROSITE; PS01999; ZF-RANBP2_2; 1.
 CC Nuclear protein; zinc-finger; Metal-binding; RNA-binding;
 CC Alternative splicing.
 CC DOMAIN 12 111 GLY-RICH.
 CC DOMAIN 119 205 RNA-BINDING (RRM).
 CC DOMAIN 212 275 GLY-RICH.
 CC ZN_FING 280 309 RANBP2-TYPE.
 CC DOMAIN 312 391 GLY-RICH.
 CC VARSPLIC 4 47 MISSING (IN ISOFORM 1).
 CC CONFLICT 39 41 PNY -> LFI (IN REF. 4).
 CC CONFLICT 92 92 P -> H (IN REF. 3).
 CC CONFLICT 108 108 G -> GG (IN REF. 3).
 CC CONFLICT 253 258 MISSING (IN REF. 3).
 CC CONFLICT 283 283 D -> E (IN REF. 4 AND 5).
 CC CONFLICT 389 398 DGGPMRNDGG -> WVDQEKRS (IN REF. 4).
 CC SEQUENCE 404 AA; 39141 MW; 7062A0446BEA5984 CRC64;
 SQ
 Query Match 31.9%; Score 71.5; DB 1; Length 404;
 Best Local Similarity 59.3%; Pred. No. 2.8;
 Matches 16; Conservative 0; Mismatches 8; Indels 3; Gaps 1;
 QY 1 GGGGGIEGFTLRQWLARAGGGGGGG 27
 DB 214 GGGGGGGG---RGGFGRGGGGGGGG 237
 RESULT 15
 SR54_ASPNG STANDARD; PRT; 534 AA.
 ID SR54_ASPNG
 AC Q00179;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Signal recognition particle 54 kDa protein homolog.
 OS *Aspergillus niger*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]

DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 FT Transcription regulation.
 FT DOMAIN 129 134 ANTP-TYPE HEXAPEPTIDE.
 FT DOMAIN 154 181 GLY-RICH.
 FT DNA_BIND 191 250 HOMEBOX.
 FT CONFLICT 113 113 G -> C (IN REF. 1).
 FT CONFLICT 119 119 A -> S (IN REF. 1).
 FT CONFLICT 152 169 GCGGGGGGGGGGGGGG -> RLWWRPVAVAAAAAARG
 (IN REF. 3).
 FT CONFLICT 182 182 D -> N (IN REF. 4).
 FT CONFLICT 216 217 LC -> FV (IN REF. 3).
 FT CONFLICT 330 330 S -> L (IN REF. 3).
 FT CONFLICT 342 361 GAYGTMTGSPVYVGGGY -> APTGRPPCRAVRCMWAG
 VAT (IN REF. 3).
 SQ SEQUENCE 433 AA; 44353 MW; 9AD3C922663612A6 CRC64;
 Query Match 32.1%; Score 72; DB 1; Length 433;
 Best Local Similarity 48.4%; Pred. No. 2.7;
 Matches 15; Conservative 1; Mismatches 5; Indels 10; Gaps 1;
 QY 1 GGGGTEGPTLRQWLAAARAGGGGGGIEGP 31
 ||||| | ||||| | |
 Db 165 GGGGGSG-----GGGGGGGDKSP 185
 RESULT 13
 OVO_DROME
 ID OVO_DROME STANDARD; PRT; 1028 AA.
 AC P5121; O9XZU4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ovo protein (Shaven baby protein).
 GN OVO OR SVB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
 melanogaster: relationship to genetic complexity.";
 RL Mol. Cell. Biol. 14:6809-6818(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE=91293102; PubMed=1712294;
 RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
 RT "The ovo gene of Drosophila encodes a zinc finger protein required
 for female germ line development.";
 RL EMOB J. 10:2259-2266(1991).
 CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND
 ACCUMULATES IN NURSE CELLS DURING OOCYTESIS. STORED IN THE EGG,
 BUT IS RAPIDLY LOST IN THE EMBRYO EXCEPT FOR ITS CONTINUED
 PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U11383; AAB60216.1; -

DR EMBL; X59772; CAB36921.1; ALT_SEQ.
 DR HSP; P25490; 12NM.
 DR FlyBase; FBgn0003028; ovo.
 DR InterPro; IPR000822; Znf-C2H2.
 DR Pfam; PF00096; zf-C2H2; 4.
 DR PRINTS; PR00048; ZINCFINGER.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 Transcription regulation.
 FT DOMAIN 62 66 POLY-ALA.
 FT DOMAIN 72 77 POLY-GLY.
 FT DOMAIN 80 85 POLY-GLY.
 FT DOMAIN 98 108 POLY-GLY.
 FT DOMAIN 144 152 POLY-HIS.
 FT DOMAIN 153 159 POLY-ASN.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 353 POLY-GLN.
 FT DOMAIN 357 361 POLY-GLN.
 FT DOMAIN 410 414 POLY-GLN.
 FT DOMAIN 418 422 POLY-GLN.
 FT DOMAIN 426 432 POLY-GLN.
 FT DOMAIN 445 453 POLY-GLN.
 FT DOMAIN 456 459 POLY-GLN.
 FT DOMAIN 466 474 POLY-GLN.
 FT DOMAIN 497 517 POLY-ALA.
 FT DOMAIN 524 529 POLY-SER.
 FT DOMAIN 549 558 POLY-ALA.
 FT DOMAIN 639 651 POLY-ALA.
 FT DOMAIN 717 725 POLY-ALA.
 FT DOMAIN 797 802 POLY-GLN.
 FT DOMAIN 820 823 POLY-GLN.
 FT DOMAIN 826 832 POLY-GLN.
 FT DOMAIN 874 992 ZINC FINGERS.
 FT ZN_FING 874 896 C2H2-TYPE.
 FT ZN_FING 902 924 C2H2-TYPE.
 FT ZN_FING 930 953 C2H2-TYPE.
 FT ZN_FING 969 992 C2H2-TYPE.
 FT CONFLICT 647 647 A -> R (IN REF. 2).
 SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;
 Query Match 32.1%; Score 72; DB 1; Length 1028;
 Best Local Similarity 55.2%; Pred. No. 5.7;
 Matches 16; Conservative 2; Mismatches 9; Indels 2; Gaps 1;
 QY 2 GGGGIEGPTLRQWLAAARAGGGGGGIEG 30
 |||| | : ||||| |
 Db 82 GGGGASGP--GGGPSANSGGGGGGGNG 108
 RESULT 14
 CAZ_DROME
 ID CAZ_DROME STANDARD; PRT; 404 AA.
 AC Q27294; Q24445; Q9VXI4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE RNA-binding protein cabeza (Sarcoma-associated RNA-binding fly
 homolog) (P19).
 GN CAZ OR SARFH OR CG3606.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN-CANTON-S;
 RX MEDLINE=95349623; PubMed=7623847;
 RA Immanuel D., Zinszner H., Ron D.;
 RT "Association of SARFH (sarcoma-associated RNA-binding fly homolog)
 with regions of chromatin transcribed by RNA polymerase II.";

```

RP SEQUENCE FROM N.A.
RC TISSUE=Skin, and Uterus;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE-20105516; PubMed-10639123;
RA Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H.,
RA Nakagawa K., Irie A., Sorimachi H., Dourenkow G., Bartunik H.,
RA Suzuki K., Bode W.;
RT "The crystal structure of calcium-free human m-calpain suggests an
RT electrostatic switch mechanism for activation by calcium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodeling and signal transduction.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding (By similarity).
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X04106; CAA27726.1; -
CC EMBL; M31511; AAA35646.1; -
CC EMBL; M31502; AAA35646.1; JOINED.
CC EMBL; M31503; AAA35646.1; JOINED.
CC EMBL; M31504; AAA35646.1; JOINED.
CC EMBL; M31505; AAA35646.1; JOINED.
CC EMBL; M31506; AAA35646.1; JOINED.
CC EMBL; M31507; AAA35646.1; JOINED.
CC EMBL; M31508; AAA35646.1; JOINED.
CC EMBL; M31509; AAA35646.1; JOINED.
CC EMBL; M31510; AAA35646.1; JOINED.
CC EMBL; AD001527; AAB51183.1; -
CC EMBL; AC002984; AAB81546.1; -
CC EMBL; BC000592; AAB00592.1; -
CC EMBL; BC007779; AAB07779.1; -
CC PIR; A23650; CIHUL.
CC PDB; 1KFU; 07-DEC-01.
CC MIM; 114170; -
CC InterPro: IPR002048; EF-hand.
CC Pfam; PF00036; ehand; 3
CC PROSITE; PS00018; EF_HAND; 2.
CC Calcium-binding; Repeat; 3D-structure.
CC DOMAIN 1 66 GLY-RICH (HYDROPHOBIC).
CC FT DOMAIN 96 268 CALCIUM-BINDING.
CC FT CA_BIND 152 163 EF-HAND 1.
CC FT CA_BIND 182 193 EF-HAND 2.
CC FT DOMAIN 217 228 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
CC FT DOMAIN 247 260 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
CC FT DOMAIN 10 26 POLY-GLY.
CC FT DOMAIN 35 56 POLY-GLY.
CC FT DOMAIN 78 83 POLY-PRO.
CC SEQUENCE 268 AA; 28316 MW; 17B87A8E47A90632 CRC64;
SQ
Query Match 32.1%; Score 72; DB 1; Length 268;
Best Local Similarity 55.2%; Pred. No. 1.8;
Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;
OY 1 GGGGGIEG--PTLRQWLAARAGGGGGGG 27
DB 16 GGGGGLGGGLGNVLGGISGAGGGGGG 44
RESULT 12

```

```

HXB3_MOUSE
ID HXB3_MOUSE STANDARD; PRT; 433 AA.
AC P09026; P10285; Q61680;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Homeobox protein Hox-B3 (Hox-2.7) (MH-23).
GN HOXB3 OR HOXB-3 OR HOX-2.7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92258392; PubMed-1582411;
RA Sham M.H., Hunt P., Nonchev S., Papalopolu N., Graham A.,
RA Boncinelli E., Krumlauf R.;
RT "Analysis of the murine Hox-2.7 gene: conserved alternative
RT transcripts with differential distributions in the nervous system and
RT the potential for shared regulatory regions."
RL EMBO J. 11:1825-1836(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-95196953; PubMed-7890121;
RA Brown W.M., Taylor G.R.;
RT "The 5'-sequence of the murine Hox-b3 (Hox-2.7) gene and its intron
RT contain multiple transcription-regulatory elements."
RL Int. J. Biochem. 26:1403-1409(1994).
RN [3]
RP SEQUENCE OF 152-361 FROM N.A.
RX MEDLINE-88054465; PubMed-2890503;
RA Lonai P., Arman E., Czosnek H., Ruddle F.H., Blatt C.;
RT "New murine homeoboxes: structure, chromosomal assignment, and
RT differential expression in adult erythropoiesis."
RL DNA 6:409-418(1987).
RN [4]
RP SEQUENCE OF 181-265 FROM N.A.
RX MEDLINE-89091992; PubMed-2463210;
RA Graham A., Papalopolu N., Lorimer J., McVey J.H., Tuddenham E.G.D.,
RA Krumlauf R.;
RT "Characterization of a murine homeo box gene, Hox-2.6, related to the
RT Drosophila Deformed gene."
RL Genes Dev. 2:1424-1438(1988).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X66177; CAA46951.1; -
CC EMBL; U02278; AAB60496.1; -
CC EMBL; M18168; AAA37840.1; -
CC PIR; S20963; S20963.
CC PIR; C29585; C29585.
CC HSPF; P02833; ISAN.
CC TRANSFAC; TUI724; -
CC MGD; MGI:96184; Hoxb3.
CC InterPro; IPR001827; Antennapedia.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00025; ANTENNAPEDIA.
CC PRINTS; PR00024; HOMEBOX.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00032; ANTENNAPEDIA; 1.

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M11778; AAA31010.1; -
 DR EMBL; M11779; AAA31011.1; -
 DR PIR; A25166; CIPGL.
 DR PDB; 1ALV; 03-JUN-98.
 DR PDB; 1ALW; 10-JUN-98.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 KW Calcium-binding; Repeat; Acetylation; 3D-structure.
 FT MOD_RES 1 1
 FT DOMAIN 1 67
 FT CA_BIND 94 266
 FT CA_BIND 150 161
 FT CA_BIND 180 191
 FT DOMAIN 215 226
 FT DOMAIN 245 258
 FT DOMAIN 10 26
 FT DOMAIN 35 54
 FT DOMAIN 76 81
 FT POLY-PRO.
 SQ SEQUENCE 266 AA; 28068 MW; 3FA81023EDC4141A CRC64;

Query Match 32.1%; Score 72; DB 1; Length 266;
 Best Local Similarity 55.2%; Pred. No. 1.8;
 Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 GGGGGIEG--PTLRQWLAAARAGGGGGGGG 27
 |||||: | : | : |||||
 Db 16 GGGGGGLGGGLGNVLGLISGAGGGGGGG 44

RESULT 10

CANS_RABIT
 ID CANS_RABIT STANDARD; PRT; 266 AA.
 AC P06813;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
 DE (Calcium-activated neutral proteinase) (CANP).
 GN CAPN4.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86250903; PubMed=3013892;
 RA Emori Y., Kawasaki H., Imajoh S., Kawashima S., Suzuki K.;
 RT "Isolation and sequence analysis of cDNA clones for the small subunit
 RT of rabbit calcium-dependent protease."
 RL J. Biol. Chem. 261:9472-9476(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87279982; PubMed=3038855;
 RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;
 RT "E-F hand structure-domain of calcium-activated neutral protease
 RT (CANP) can bind Ca²⁺ ions."
 RL J. Biochem. 101:889-895(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=101889-895(1987).
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca²⁺ binding (By similarity).
 CC PTM: The N-terminus is blocked.

CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M13364; AAA81565.1; -
 DR PIR; A24816; CIRBL.
 DR HSP; P04632; 1DKV.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 KW Calcium-binding; Repeat.
 FT DOMAIN 1 64
 FT CA_BIND 94 266
 FT CA_BIND 150 161
 FT CA_BIND 180 191
 FT DOMAIN 215 226
 FT DOMAIN 245 258
 FT DOMAIN 10 25
 FT DOMAIN 34 54
 FT DOMAIN 76 81
 FT POLY-PRO.
 SQ SEQUENCE 266 AA; 28239 MW; 1D7FE31989F70B03 CRC64;

Query Match 32.1%; Score 72; DB 1; Length 266;
 Best Local Similarity 55.2%; Pred. No. 1.8;
 Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 GGGGGIEG--PTLRQWLAAARAGGGGGGGG 27
 |||||: | : | : |||||
 Db 15 GGGGGGLGGGLGNVLGLISGAGGGGGGG 43

RESULT 11

CANS_HUMAN
 ID CANS_HUMAN STANDARD; PRT; 268 AA.
 AC P04632;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
 DE (Calcium-activated neutral proteinase) (CANP).
 GN CAPN4 OR CAPNS1 OR CAPNS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86286563; PubMed=3016651;
 RA Ohno S., Emori Y., Suzuki K.;
 RT "Nucleotide sequence of a cDNA coding for the small subunit of human
 RT calcium-dependent protease."
 RL Nucleic Acids Res. 14:5559-5559(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87066759; PubMed=3024120;
 RA Miyake S., Emori Y., Suzuki K.;
 RT "Gene organization of the small subunit of human calcium-activated
 RT neutral protease."
 RL Nucleic Acids Res. 14:8805-8817(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Gaines J.,
 RA Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,
 RA Kobayashi A., Olsen A.O., Carrano A.V.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; J05065; AAA30422.1; -
CC PIR; A34466; A34466.
CC HSP; P04574; IALV.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF00036; EF-hand; 3.
CC PROSITE; PS00018; EF_HAND; 2.
CC KEGG; Calcium-binding; Repeat.
CC FT DOMAIN 1 64
CC FT CA_BIND 91 263
CC FT CA_BIND 147 158
CC FT CA_BIND 177 188
CC FT DOMAIN 212 223
CC FT DOMAIN 242 255
CC FT DOMAIN 10 25
CC FT DOMAIN 34 51
CC SQ SEQUENCE 263 AA; 27931 MW; 9427925D5284CE1A CRC64;

Query Match 32.1%; Score 72; DB 1; Length 263;
Best Local Similarity 55.2%; Pred. No. 1.8;
Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 GGGGGIEG--PTLRQLAARAGGGGGGG 27
DB 15 GGGGGIEG--PTLRQLAARAGGGGGGG 43

RESULT 9

ID CANS_PIG STANDARD; PRT; 266 AA.
AC P04574;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
DE (Calcium-activated neutral proteinase) (CANP).
GN CAPN4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85298299; PubMed=2994060;
RA Sakihama T., Kakidani H., Zenita K., Yumoto N., Kikuchi T., Sasaki T.,
RA Kannagi R., Nakanishi S., Ohmori M., Takio K., Titani K., Murachi T.;
RT "A putative Ca²⁺-binding protein: structure of the light subunit of
RT porcine calpain elucidated by molecular cloning and protein sequence
RT analysis";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6075-6079(1985).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 94-266.
RX MEDLINE=97372891; PubMed=9228946;
RA Lin G.D., Chattopadhyay D., Maki M., Wang K.K., Carson M., Jin L.,
RA Yuen P.W., Takano E., Hatanaka M., Delucas L.J., Nardana S.V.;
RT "Crystal structure of calpain bound domain VI of calpain at 1.9-A
RT resolution and its role in enzyme assembly, regulation, and inhibitor
RT binding";
RL Nat. Struct. Biol. 4:539-547(1997).
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodeling and signal transduction.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca²⁺ binding (By similarity).
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC DETERMINING THE PHYSICAL PROPERTIES OF CUTICLES.
CC -!- TISSUE SPECIFICITY: EPIDERMAL REGIONS SYNTHESIZING HARD CUTICLE.
CC -!- DEVELOPMENTAL STAGE: PRECYDSTAL ADULT CUTICLE DEPOSITION.
CC -!- INDUCTION: BY 20-HYDROXYCYDSONE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X60455; CAA42985.1; -
CC PIR; X72783; CAA51290.1; -
CC PIR; S16063; S16063.
CC InterPro: IPR000618; Insect cuticle.
CC Pfam: PF00379; Insect cuticle; 1.
CC Structural protein; Cuticle; Signal.
CC FT SIGNAL 1 19
CC FT CHAIN 20 199
CC FT DOMAIN 21 26
CC FT DOMAIN 34 37
CC FT DOMAIN 63 72
CC FT DOMAIN 88 104
CC FT DOMAIN 138 138
CC FT CONFLICT 138 138 S -> Y (IN REF. 2).
CC SQ SEQUENCE 199 AA; 20710 MW; 1EBF2D1DBEC6739F CRC64;

Query Match 32.1%; Score 72; DB 1; Length 199;
Best Local Similarity 48.6%; Pred. No. 1.4;
Matches 17; Conservative 1; Mismatches 9; Indels 8; Gaps 1;

QY 1 GGGGGIEGPTLRQW-----LAARAGGGGGGGG 27
DB 65 GGGGGIEGPTLRQW-----LAARAGGGGGGGG 99

RESULT 8

ID CANS_BOVIN STANDARD; PRT; 263 AA.
AC P13135;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
DE (Calcium-activated neutral proteinase) (CANP).
GN CAPN4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90008911; PubMed=2551902;
RA McClelland P., Lash J.A., Hathaway D.R.;
RT "Identification of major autolytic cleavage sites in the regulatory
RT subunit of vascular calpain II. A comparison of partial
RT amino-terminal sequences to deduced sequence from complementary
RT DNA";
RL J. Biol. Chem. 264:17428-17431(1989).
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodeling and signal transduction.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca²⁺ binding (By similarity).
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

DE (HHARI) (H7-AP2) (HUSY-27) (MOP-6).
GN ARIH1 OR ARI OR UBCH7BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=99452997; PubMed=10521492;
RA Moynihan T.P., Ardley H.C., Nuber U., Rose S.A., Jones P.F.,
RA Markham A.F., Scheffner M., Robinson P.A.;
RT "The ubiquitin-conjugating enzymes Ubch7 and Ubch8 interact with RING
RT finger/IBR motif-containing domains of HHARI and H7-AP1";
RL J. Biol. Chem. 274:30963-30968(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX Trockenbacher A., Marksteiner R., Schneider R.;
RT "Human ariadne homolog";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 95-557 FROM N.A.
RX MEDLINE=20341325; PubMed=10880484;
RA Aquilera M., Oliveros M., Martinez-Padron M., Barbas J.A., Ferrus A.;
RT "Ariadne-1: a vital Drosophila gene is required in development and
RT defines a new conserved family of ring-finger proteins.";
RL Genetics 155:1231-1244(2000).
RN [4]
RP SEQUENCE OF 298-557 FROM N.A.
RX TISSUE=Brain;
RA Stanchi F., Toppo S., Dioguardi R., Simionati B.,
RA Cannata N., Zimbello R., Lanfranchi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RT yeast sequences";
RL Yeast 18:69-80(2001).
RN [5]
RP SEQUENCE OF 377-557 FROM N.A.
RX TISSUE=Monocytes;
RA Fujii Y., Takayama K., Ukai Y., Yoshimoto M.;
RT "Molecular and biological characterization of a new ring finger
RT protein, MOP-6 which is highly expressed in activated human
RT monocytes";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP INTERACTION WITH UBE2L1, AND MUTAGENESIS OF GLN-187; ILE-188; CYS-
RP 208 AND TYR-258.
RX MEDLINE=21276469; PubMed=11278816;
RA Ardley H.C., Tan N.G.S., Rose S.A., Markham A.F., Robinson P.A.;
RT "Features of the parkin/ariadne-like ubiquitin ligase, HHARI, that
RT regulate its interaction with the ubiquitin-conjugating enzyme,
RT Ubch7";
RL J. Biol. Chem. 276:19640-19647(2001).
CC -!- FUNCTION: MIGHT ACT AS AN E3 UBIQUITIN-PROTEIN LIGASE, OR AS PART
CC OF THE E3 COMPLEX, WHICH ACCEPTS UBIQUITIN FROM SPECIFIC E2
CC UBIQUITIN-CONJUGATING ENZYMES, SUCH AS UBE2L1/UBE2L3/UBCM4, AND
CC THEN TRANSFERS IT TO SUBSTRATES.
CC -!- SUBUNIT: INTERACTS WITH UBE2L1.
CC -!- SUBCELLULAR LOCATION: MAINLY CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- SIMILARITY: CONTAINS 2 RING-TYPE ZINC FINGERS.
CC -!- SIMILARITY: CONTAINS 1 IBR-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AJ243190; CAB45870.1; "

DR EMBL; AF072832; AAD28088.1; -;
DR EMBL; AJ130976; CAA10274.1; -;
DR EMBL; AJ009771; CAA08817.1; -;
DR EMBL; AB014774; BAB19786.1; -;
DR MIM: 605624; -;
DR InterPro: IPR002867; IBR.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF01485; IBR; 1.
DR SMART; SM00184; RING; 2.
DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Ubiquitin conjugation; Zinc-finger; Repeat; Coiled coil.
FT DOMAIN 10 38 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 67 92 GLY-RICH.
FT ZN_FING 186 235 RING-TYPE 1.
FT ZN_FING 256 317 RING-TYPE.
FT ZN_FING 344 389 RING-TYPE 2.
FT DOMAIN 433 449 COILED COIL (POTENTIAL).
FT DOMAIN 186 254 INTERACTION WITH UBE2L1.
FT MUTAGEN 187 188 Q1->HV: NO LOSS OF INTERACTION WITH
FT UBE2L1.
FT MUTAGEN 188 188 I->A: LOSS OF INTERACTION WITH UBE2L1.
FT MUTAGEN 208 208 C->A,H: LOSS OF INTERACTION WITH UBE2L1.
FT MUTAGEN 258 258 Y->A: NO LOSS OF INTERACTION WITH
FT UBE2L1.
FT CONFLICT 122 122 E -> D (IN REF. 2).
FT CONFLICT 227 227 H -> Q (IN REF. 2 AND 3).
FT CONFLICT 237 237 D -> N (IN REF. 3).
FT CONFLICT 303 303 F -> S (IN REF. 4).
FT CONFLICT 309 316 ENWHDVVK -> AIGMILFQ (IN REF. 4).
FT CONFLICT 322 322 K -> T (IN REF. 4).
SQ SEQUENCE 557 AA; 64126 MW; 44BCA291863ABB6A CRC64;
Query Match 32.4%; Score 72.5; DB 1; Length 557;
Best Local Similarity 51.6%; Pred. No. 3;
Matches 16; Conservative 0; Mismatches 8; Indels 7; Gaps 1;
QY 1 GGGGIEGPTLRQLAARAGGGGGGGGIEGP 31
DB 68 GGGGSALGP-----GGGGGGGGGGGGG 91
RESULT 7
AC22_TENMO
ID AC22_TENMO STANDARD; PRT; 199 AA.
AC P26968; O27015;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Adult-specific cuticular protein ACP-22 precursor.
GN ACP22.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92097825; PubMed=1728581;
RA Bouhin H., Charles J.-P., Quennedey B., Delachambre J.;
RT "Developmental profiles of epidermal mRNAs during the pupal-adult
RT molt of Tenebrio molitor and isolation of a cDNA clone encoding an
RT adult cuticular protein: effects of a juvenile hormone analogue.";
RL Dev. Biol. 149:112-122(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97242547; PubMed=9087546;
RA Bouhin H., Braquart C., Charles J.-P., Quennedey B., Delachambre J.;
RT "Nucleotide sequence of an adult-specific cuticular protein gene from
RT the beetle Tenebrio molitor: effects of 20-hydroxyecdysone on mRNA
RT accumulation";
RL Insect Mol. Biol. 2:81-88(1993).
CC -!- FUNCTION: CUTICULAR PROTEINS PLAY A SIGNIFICANT ROLE IN

Wed Oct 9 10:30:20 2002

QY 1 GGGGGIE---GPTLRQWLAAARAGGGGGG 27
 |||||: | | : |||||
 Db 11 GGGGGLGGGLGNVLGLISGAAGGGGGG 40

RESULT 4
 ID CANS_MOUSE STANDARD; PRT; 269 AA.
 AC 08456:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
 DE (Calcium-activated neutral proteinase) (CAMP).
 GN CAPNA
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20285452; PubMed=10825211;
 RA Arthur J.S.C., Elce J.S., Hegadorn C., Williams K., Greer P.A.;
 RT "Disruption of the murine calpain small subunit gene, Capn4: calpain
 RT is essential for embryonic development but not for cell growth and
 RT division.";
 RT Mol. Cell. Biol. 20:4474-4481(2000).
 RL
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC
 CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC
 CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF058298; AAC97194.1; -
 CC HSP; P04632; IDK.
 CC DR MGD; MG1:88266; Capn4.
 CC DR InterPro; IPR002048; EF-hand.
 CC DR Pfam; PF00036; efhand; 3.
 CC DR PROSITE; PS00018; EF_HAND; 2.
 CC
 CC Calcium-binding; Repeat.
 CC
 CC FT DOMAIN 1 65 GLY-RICH (HYDROPHOBIC).
 CC FT DOMAIN 97 269 CALCIUM-BINDING.
 CC FT CA_BIND 153 164 EF_HAND 1 (POTENTIAL).
 CC FT CA_BIND 183 194 EF_HAND 2 (POTENTIAL).
 CC FT CA_BIND 218 229 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 CC FT DOMAIN 248 261 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 CC SEQUENCE 269 AA; 28463 MW; C578771942FB57E9 CRC64;
 SQ

Query Match 33.3%; Score 74.5; DB 1; Length 269;
 Best Local Similarity 53.3%; Pred. No. 1.1;
 Matches 16; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

QY 1 GGGGGIE---GPTLRQWLAAARAGGGGGG 27
 |||||: | | : |||||
 Db 15 GGGGGLGGGLGNVLGLISGAAGGGGGG 44

RESULT 5
 ID SXL_CERCA STANDARD; PRT; 348 AA.
 AC 061374:
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sex-lethal protein homolog (CCSLX).
 GN SXL.
 OS Ceratitidis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritidae; Tephritidae; Ceratitidis.
 OX NCBI_TaxID=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GENAKIO;
 RX MEDLINE=98171464; PubMed=9502730;
 RA Saccone G., Pelluso I., Artiano D., Giordano E., Bopp D., Polito L.C.;
 RT "The Ceratitidis capitata homologue of the Drosophila sex-determining
 RT gene Sex-lethal is structurally conserved, but not sex-specifically
 RT regulated.";
 RT Development 125:1495-1500(1998).
 RL
 CC -!- FUNCTION: UNKNOWN; APPARENTLY NOT INVOLVED IN SOMATIC SEX
 CC DETERMINATION.
 CC
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS; ADULT-SPECIFIC ISOFORMS
 CC A1, A2, A3, A4, AND EMBRYO-SPECIFIC ISOFORMS E1, E2 AND E3 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS OF BOTH SEXES. ALSO
 CC EXPRESSED IN THE PROGENITOR CELLS OF THE GERMLINE.
 CC
 CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF026145; AAC38968.1; -
 CC HSP; P19339; ISXL.
 CC DR InterPro; IPR000504; RRM.
 CC DR Pfam; PF00076; rrm; 2.
 CC DR PRINTS; PR00961; HUDSLXRNA.
 CC DR SMART; SM00360; RRM; 2.
 CC DR PROSITE; PS0102; RRM; 2.
 CC DR PROSITE; PS00030; RRM_RNP_1; 1.
 CC
 CC RNA-binding; Repeat; Nuclear protein; Alternative splicing.
 CC
 CC FT DOMAIN 1 27 GLY/ASN-RICH DOMAIN.
 CC FT DOMAIN 110 188 RNA-BINDING (RRM) 1.
 CC FT DOMAIN 196 276 RNA-BINDING (RRM) 2.
 CC FT DOMAIN 68 75 POLY-GLY.
 CC FT DOMAIN 95 99 POLY-GLY.
 CC FT DOMAIN 293 311 POLY-GLY.
 CC FT DOMAIN 312 316 POLY-PRO.
 CC FT VARSPLIC 37 44 MISSING (IN ISOFORM AL).
 CC SEQUENCE 348 AA; 37188 MW; CABA3DA5C2C8874A CRC64;
 SQ

Query Match 33.3%; Score 74.5; DB 1; Length 348;
 Best Local Similarity 51.6%; Pred. No. 1.3;
 Matches 16; Conservative 1; Mismatches 3; Indels 11; Gaps 1;

QY 1 GGGGGIEGPTLRQWLAAARAGGGGGGIEGP 31
 |||||: | | : |||||
 Db 293 GGGGGGGG-----GGGGGGGGGGG 312

RESULT 6
 ID ARIL_HUMAN STANDARD; PRT; 557 AA.
 AC Q9V4X5; Q9UP39; Q9UEN0; O76026; Q9H3T6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Ariadne-1 protein homolog (ARI-1) (Ubiquitin-conjugating enzyme E2-
 DE binding protein 1) (UbcH7-binding protein) (UbcM4-interacting protein)

RESULT 2

BT3_MESAU STANDARD; PRT; 367 AA.
 AC 009029;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE BT33 protein.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 NCBI_TaxID=10036;
 (1)
 SEQUENCE FROM N.A.
 MEDLINE=96140430; PubMed=8552091;
 RA Peyton M., Stellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,
 RA Tsai M.-J.;
 RT "BETA3, a novel helix-loop-helix protein, can act as a negative
 RT regulator of BETA2 and MyoD-responsive genes.";
 RL Mol. Cell. Biol. 16:626-633(1996).
 CC -!- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3
 CC (E47) / NEURODI HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF
 CC NEURODI AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION
 CC WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE
 CC OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.
 CC -!- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
 CC -!- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; S80870; AAB50691.1; -;
 CC InterPro; IPR003015; HLH_MYC.
 CC InterPro; IPR001092; HLH_dlm.
 CC Pfam; PF00010; HLH; 1.
 CC SMART; SM00353; HLH; 1.
 CC DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
 CC Nuclear protein; Transcription regulation; Repressor.
 FT DOMAIN 11 14 POLY-ALA.
 FT DOMAIN 58 62 POLY-SER.
 FT DOMAIN 83 99 POLY-GLY.
 FT DOMAIN 174 179 POLY-GLY.
 FT DOMAIN 204 217 POLY-GLY.
 FT DNA_BIND 229 240 BASIC DOMAIN.
 FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 311 319 POLY-ALA.
 SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AFF96E85F77 CRC64;

Query Match 33.5%; Score 75; DB 1; Length 367;
 Best Local Similarity 48.5%; Pred. No. 1.2;
 Matches 16; Conservative 1; Mismatches 6; Indels 10; Gaps 1;

QY 1 GGGGGTGTTLROWLAARAGGGGGGGGEGPTL 33
 ||||| |
 Db 83 GGGGGAGG-----GGGGGGGGGVPGL 105

RESULT 3

CANS_RAT
 ID CANS_RAT STANDARD; PRT; 266 AA.
 AC Q64337; P97572;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
 DE (Calcium-activated neutral proteinase) (CANP) (Fragment).
 GN CAPNA OR CSS1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 (1)
 SEQUENCE FROM N.A.
 MEDLINE=97107433; PubMed=8950173;
 RA Sorimachi H., Amano S., Ishiura S., Suzuki K.;
 RT "Primary sequences of rat mu-calpain large and small subunits are,
 RT respectively, moderately and highly similar to those of human.";
 RL Biochim. Biophys. Acta 1309:37-41(1996).
 (2)
 SEQUENCE OF 83-266 FROM N.A., AND PARTIAL SEQUENCE.
 RP MEDLINE=95074051; PubMed=7982961;
 RA Graham-Siegenthaler K., Gauthier S., Davies P.L., Elce J.S.;
 RT "Active recombinant rat calpain II. Bacterially produced large and
 RT small subunits associate both in vivo and in vitro.";
 RL J. Biol. Chem. 269:30457-30460(1994).
 (3)
 X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 94-266.
 RP MEDLINE=97372890; PubMed=9228945;
 RA Blanchard H., Grochulski P., Li Y., Arthur J.S.C., Davies P.L.,
 RA Elce J.S., Cygler M.;
 RT "Structure of a calpain Ca(2+)-binding domain reveals a novel EF-hand
 RT and Ca(2+)-induced conformational changes.";
 RL Nat. Struct. Biol. 4:532-538(1997).
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodelling and signal transduction.
 CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U53859; AAC53002.1; -;
 CC EMBL; U10861; AAG4828.1; -;
 CC PDB; 1AJ5; 20-MAY-98.
 CC PDB; 1DVI; 27-MAY-98.
 CC InterPro; IPR002048; EF-hand.
 CC Pfam; PF00036; efhand; 3.
 CC PROSITE; PS00018; EF_HAND; 2.
 CC Calcium-binding; Repeat; 3D-structure.
 FT NON_TER 1 1
 FT DOMAIN <1 62 GLY-RICH (HYDROPHOBIC).
 FT DOMAIN 94 266 CALCIUM-BINDING.
 FT CA_BIND 150 161 EF-HAND 1.
 FT CA_BIND 180 191 EF-HAND 2.
 FT DOMAIN 215 226 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 245 258 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 6 21 POLY-GLY.
 FT DOMAIN 33 54 POLY-GLY.
 FT DOMAIN 76 81 POLY-PRO.
 FT CONFLICT 83 83 S -> M (IN REF. 2).
 SQ SEQUENCE 266 AA; 28079 MW; 7BF8E444576D835 CRC64;

Query Match 33.3%; Score 74.5; DB 1; Length 266;
 Best Local Similarity 53.3%; Pred. No. 1.1;
 Matches 16; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.89696 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-34
Perfect score: 224
Sequence: 1 GGGGGIEGPTLRQWLARAG.....GGGGGIEGPTLRQWLARA 41

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	34.8	323	1 JUND_CHICK	P27921 gallus gall
2	75	33.5	367	1 BET3_MESAU	O09029 mesocricetu
3	74.5	33.3	266	1 CANS_RAT	Q64537 rattus norv
4	74.5	33.3	269	1 CANS_MOUSE	O88456 mus musculu
5	74.5	33.3	348	1 SXL_CERCA	O61374 ceratitidis c
6	72.5	32.4	557	1 ARIL_HUMAN	O9V4X5 homo sapien
7	72	32.1	199	1 AC22_TENMO	P26968 tenebrio mo
8	72	32.1	263	1 CANS_BOVIN	P13135 bos taurus
9	72	32.1	266	1 CANS_PIG	P04574 sus scrofa
10	72	32.1	266	1 CANS_RABIT	P06813 oryctolagus
11	72	32.1	266	1 CANS_HUMAN	P04632 homo sapien
12	72	32.1	433	1 HXB3_MOUSE	P09026 mus musculu
13	72	32.1	1028	1 OVO_DROME	P51521 drosophila
14	71.5	31.9	404	1 CAZ_DROME	Q27294 drosophila
15	71	31.7	534	1 SR54_ASPNG	Q00179 aspergillus
16	71	31.7	1322	1 SUS_DROME	P22293 drosophila
17	70.5	31.5	431	1 HXB3_HUMAN	P14651 homo sapien
18	70	31.2	331	1 SHX2_HUMAN	O60902 homo sapien
19	70	31.2	386	1 RB87_DROME	P48810 drosophila
20	69.5	31.0	369	1 TMAF_AVIS4	P23091 avian muscu
21	69	30.8	342	1 ROAL_SCHAM	P21522 schistocerc
22	69	30.8	384	1 GRP1_PETHY	P09789 petunia hyb
23	69	30.8	1250	1 TP3A_DROME	O9ng98 drosophila
24	68.5	30.6	369	1 MAF2_RAT	P54844 rattus norv
25	68.5	30.6	520	1 WASP_MOUSE	P70315 mus musculu
26	68.5	30.6	619	1 LACL_NEUCR	P06811 neurospora
27	68.5	30.6	619	1 LACL2_NEUCR	P10574 neurospora
28	68.5	30.6	757	1 CIKE_HUMAN	Q14003 homo sapien
29	68.5	30.6	759	1 TOP3_CABEL	O61660 caenorhabdi
30	68.5	30.6	769	1 CIKE_MOUSE	Q63959 mus musculu
31	68.5	30.6	779	1 M130_STRPU	P08472 strongyloce
32	68.5	30.6	889	1 CIKE_RAT	Q01956 rattus norv
33	68.5	30.6	1901	1 YZ08_MYCTU	O53553 mycobacteri

34	68	30.4	331	1 SHX2_MOUSE	P70390 mus musculu
35	68	30.4	421	1 BR3A_MOUSE	P17208 mus musculu
36	68	30.4	423	1 BR3A_HUMAN	Q01851 homo sapien
37	68	30.4	518	1 FUS_MOUSE	P56959 mus musculu
38	68	30.4	958	1 TRH_DROME	Q24119 drosophila
39	67.5	30.1	141	1 YORE_TTV1	P19289 thermoprote
40	67.5	30.1	268	1 EP34_HCMVA	P16768 human cytom
41	67.5	30.1	338	1 GRP_ARATH	P27483 arabidopsis
42	67.5	30.1	331	1 SOX1_MOUSE	P53783 mus musculu
43	67.5	30.1	503	1 WAP1_HUMAN	O43516 homo sapien
44	67.5	30.1	647	1 COAT_ADVG	P24029 aleutian mi
45	67.5	30.1	684	1 EP84_HCMVA	P17151 human cytom

ALIGNMENTS

RESULT 1
JUND_CHICK STANDARD; PRT; 323 AA.
AC P27921;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transcription factor jun-D.
GN JUND.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97019832; PubMed=1923529;
RA Hartl M., Hutchins J.T., Vogt P.K.;
RT "The chicken jun gene and its product.";
CC -!- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. JUN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; X60063; CAA42665.1; -
PIR; S20099; S20099.
HSSP; P05412; IFOS.
DR TRANSFAC; T02196; -
DR InterPro; IPR002112; Leuzip_Jun.
DR InterPro; IPR001871; bZIP.
Pfam; PF00170; bZIP; 1.
PRINTS; PR00043; LEUZIPPJUN.
SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR Transcription regulation; DNA-binding; Activator; Nuclear protein.
KW DOMAIN 59 67
FT DOMAIN 155 166 POLY-GLY.
FT DNA_BIND 242 266 BASIC MOTIF.
FT DOMAIN 270 298 LEUCINE-ZIPPER.
SQ SEQUENCE 323 AA; 33205 MW; A7F6D21A97DBB676 CRC64;

Query Match 34.8%; Score 78; DB 1; Length 323;
Best Local Similarity 57.6%; Pred. No. 0.59;
Matches 19; Conservative 0; Mismatches 8; Indels 6; Gaps 1;
QY 1 GGGGGIEGPTLRQWLARAGGGGGGGGGGGGPTL 33
||||| | | | | | | | | | | | | | | | | |
Db 142 GGGGGPNGG-----AAAAGGGGGGGGGGGGEL 168

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987
 A: Accession: H70839
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-496 <COL>
 A: Cross-references: GB:AL021926; GB:AL123456; NID: g3261520; PIDN: CAA17303.1; PID: g290957
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: Rv0109
 C: Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 31.7%; Score 71; DB 2; Length 496;
 Best Local Similarity 50.0%; Pred. No. 3.9;
 Matches 14; Conservative 3; Mismatches 7; Indels 4; Gaps 1;

QY 1 GGGGGTGGPTLRWLAAAGGGGGGGI 28
 || ||: || : || || || ||:
 Db 173 GSGGVGGP----GIAGSAGAGGAGGL 196

Search completed: October 9, 2002, 09:05:13
 Job time: 10.2178 secs

us-09-422-838c-34.rpr

Wed Oct 9 10:30:19 2002

```
A:Gene: CESP:Y50E8A.g
A:Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match      31.9%; Score 71.5; DB 2; Length 1585;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 16; Conservative 2; Mismatches 9; Indels 5; Gaps 1;

Qy 1 GGGGIEGPTLRQWLAAARAGGGGGGIEGPT 32
    ||||| | | | | | | | | | | | |
Db 462 GGGGGAGG-----GYAKPSGGGGGGYAKPS 488

RESULT 28
Tl3021
hypothetical protein F8L21.90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 22-Oct-1999
C:Accession: Tl3021
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;
submitted to the Protein Sequence Database, July 1999
A:Reference number: 217587
A:Accession: Tl3021
A:Molecule type: DNA
A:Residues: 1-371 <BEV>
A:Cross-references: EMBL:AL096882; GSPDB:GN00062; ATSP:F8L21.90
A:Experimental source: cultivar Columbia; BAC clone F8L21
C:Genetics:
A:Gene: ATSP:F8L21.90
A:Map position: 4

Query Match      31.7%; Score 71; DB 2; Length 371;
Best Local Similarity 44.2%; Pred. No. 3;
Matches 19; Conservative 3; Mismatches 17; Indels 4; Gaps 3;

Qy 1 GGGGIEGPTLRQWLAAARAGG-GGGGGIEGPTLRQWLAAAR 40
    ||||| | | | | | | | | | | | |
Db 173 GGGGRRST-DQWSSFGRRSGGGGGGGAGCVSKNWSAAK 214

RESULT 29
T09084
phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)
C:Species: Chlamydomonas reinhardtii
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09084
R:Molendijk, A.J.; Irvine, R.F.
Plant Mol. Biol. 37, 53-66, 1998
A:Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinol
A:Reference number: Z16411; MUID:98281574
A:Accession: T09084
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-490 <MOL>
A:Cross-references: EMBL:U97663; NID:g2109290; PIDN:AAC50018.1; PID:g2109291
A:Experimental source: strain cw-15
C:Genetics:
A:Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match      31.7%; Score 71; DB 2; Length 490;
Best Local Similarity 47.6%; Pred. No. 3.8;
Matches 20; Conservative 2; Mismatches 10; Indels 10; Gaps 3;

Qy 1 GGGGIEGPTLRQWLAAARAGGGGGGGI---EGPTLR--OWL 37
    ||||| | | | | | | | | | | | |
Db 224 GIGSGPLGP-----LLAAGGGGGGGGGSPDGSTARDEWL 260

RESULT 30
H70839
hypothetical glycine-rich protein Rv0109 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70839
```

```
A:Accession: S54729
A:Molecule type: DNA
A:Residues: 1-404 <STO>
A:Cross-references: EMBL:L37083; NID:g567105; PIDN:AAC41563.1; PID:g567106
R:Stolow, D.T.; Haynes, S.R.
Nucleic Acids Res. 23, 835-843, 1995
A:Title: Cabeza, a drosophila gene encoding a novel RNA binding protein, shares homology
A:Reference number: S54728; MUID:95223793
A:Accession: S54728
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 118-203; 273-310 <STW>
A:Cross-references: EMBL:L37083
C:Genetics:
A:Gene: cabeza
A:Cross-references: FlyBase:FBgn0011571
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
C:Keywords: RNA binding
F:120-195/Domain: ribonucleoprotein repeat homology <RRM>

Query Match      31.9%; Score 71.5; DB 2; Length 404;
Best Local Similarity 59.3%; Pred. No. 2.9;
Matches 16; Conservative 0; Mismatches 8; Indels 3; Gaps 1;

Qy 1 GGGGIEGPTLRQWLAAARAGGGGGGG 27
    ||||| | | | | | | | | | | | |
Db 214 GGGGGGGG---RGFGRRGGGGGGGG 237

RESULT 26
Tl3855
suppressor of sable protein homolog - fruit fly (Drosophila virilis)
C:Species: Drosophila virilis
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001
C:Accession: Tl3855
R:Steinbauer, W.K.; Sterling, J.F.; Graves, J.P.
submitted to the EMBL Data Library, February 1995
A:Description: Comparison of suppressor of sable [su(s)] in two Drosophila species' reveal
A:Reference number: Z14224
A:Accession: Tl3855
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1473 <ST>
A:Cross-references: EMBL:U20660; NID:g671707; PID:g671708; PIDN:AAA62307.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0013921
A:Introns: 112/1; 216/3; 702/2; 735/1

Query Match      31.9%; Score 71.5; DB 2; Length 1473;
Best Local Similarity 53.3%; Pred. No. 9.1;
Matches 16; Conservative 0; Mismatches 3; Indels 11; Gaps 1;

Qy 1 GGGGIEGPTLRQWLAAARAGGGGGGIEG 30
    ||||| | | | | | | | | | | | |
Db 342 GGGGGAGG-----GGGGGGGGICG 360

RESULT 27
T31611
hypothetical protein Y50E8A.g - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31611
R:Steward, C.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21047
A:Accession: T31611
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1585 <WIL>
A:Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55050.1; CESP:Y50E8A.g
A:Experimental source: clone Y50E8A
C:Genetics:
```


us-09-422-838c-34.rpr

Wed Oct 9 10:30:19 2002

```

Query Match      32.1%; Score 72; DB 1; Length 266;
Best Local Similarity 55.2%; Pred. No. 1.8;
Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 GGGGGIEG--PTLRQWLAARAGGGGGGG 27
      |||||: | : | : |||||
Db 16 GGGGGLGGGLGNVLGGLISGAGGGGGG 44

RESULT 17
CIRBL
calpain (EC 3.4.22.17) small chain - rabbit
N:Alternate names: calcium-activated neutral proteinase (CANP); calpain light chain; cal
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: A24816
R:Emori, Y.; Kawasaki, H.; Imaioh, S.; Kawashima, S.; Suzuki, K.
J. Biol. Chem. 261, 9472-9476, 1986
A:Title: Isolation and sequence analysis of cDNA clones for the small subunit of rabbit
A:Reference number: A24816; MUID:86250903
A:Accession: A24816
A:Molecule type: mRNA
A:Residues: 1-266 <EMO>
A:Cross-references: GB:M13364; NID:q164875; PIDN:AAA81565.1; PID:q164876
C:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: Cleaves preferentially after tyrosine, methionine, or arginine residues and bef
C:Superfamily: calpain small chain; calmodulin repeat homology
C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer; hyd
F:1-54/Domain: glycine-rich <GLY>
F:94-125/Domain: calmodulin repeat homology <EF1>
F:137-169/Domain: calmodulin repeat homology <EF2>
F:170-199/Domain: calmodulin repeat homology <EF3>
F:202-234/Domain: calmodulin repeat homology <EF4>
F:235-266/Domain: calmodulin repeat homology <EF5>

Query Match      32.1%; Score 72; DB 1; Length 266;
Best Local Similarity 55.2%; Pred. No. 1.8;
Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 GGGGGIEG--PTLRQWLAARAGGGGGGG 27
      |||||: | : | : |||||
Db 16 GGGGGLGGGLGNVLGGLISGAGGGGGG 44

RESULT 18
CIHUL
calpain (EC 3.4.22.17) small chain - human
N:Alternate names: calcium-activated neutral proteinase (CANP)
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: A26107; A23650
R:Miyaoka, S.; Emori, Y.; Suzuki, K.
Nucleic Acids Res. 14, 8805-8817, 1986
A:Title: Gene organization of the small subunit of human calcium-activated neutral prote
A:Reference number: A93648; MUID:87066759
A:Accession: A26107
A:Molecule type: DNA
A:Residues: 1-268 <MIY>
A:Cross-references: GB:M31502
R:Ohno, S.; Emori, Y.; Suzuki, K.
Nucleic Acids Res. 14, 5559, 1986
A:Title: Nucleotide sequence of a cDNA coding for the small subunit of human calcium-dep
A:Reference number: A93631; MUID:86286563
A:Accession: A23650
A:Molecule type: mRNA
A:Residues: 1-268 <OHN>
A:Cross-references: EMBL:X04106; NID:g35327; PIDN:CAA27726.1; PID:g35328
C:Genetics:
A:Gene: GDB:CAPN4
A:Cross-references: GDB:119752; OMIM:114170

```

```

A:Map position: 19pter-19qter
A:Introns: 70/2; 81/3; 111/3; 131/1; 152/3; 175/3; 202/1; 241/1; 260/3
C:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: Cleaves preferentially after tyrosine, methionine, or arginine residues and b
C:Superfamily: calpain small chain; calmodulin repeat homology
C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer;
F:1-56/Domain: glycine-rich <GLY>
F:96-127/Domain: calmodulin repeat homology <EF1>
F:139-171/Domain: calmodulin repeat homology <EF2>
F:172-201/Domain: calmodulin repeat homology <EF3>
F:204-236/Domain: calmodulin repeat homology <EF4>
F:237-268/Domain: calmodulin repeat homology <EF5>

Query Match      32.1%; Score 72; DB 1; Length 268;
Best Local Similarity 55.2%; Pred. No. 1.8;
Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 GGGGGIEG--PTLRQWLAARAGGGGGGG 27
      |||||: | : | : |||||
Db 16 GGGGGLGGGLGNVLGGLISGAGGGGGG 44

RESULT 19
T23416
hypothetical protein K07F5.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T23416
R:Hembry, C.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19738
A:Accession: T23416
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-290 <HIL>
A:Cross-references: EMBL:Z70284; PIDN:CAA94280.1; GSPDB:GN00022; CESP:K07F5.11
A:Experimental source: clone K07F5
C:Genetics:
A:Gene: CESP:K07F5.11
A:Map position: 4
A:Introns: 89/1; 257/1
C:Superfamily: Phaseolus glycin-rich cell wall protein 1.8

Query Match      32.1%; Score 72; DB 2; Length 290;
Best Local Similarity 59.3%; Pred. No. 1.9;
Matches 16; Conservative 1; Mismatches 6; Indels 4; Gaps 1;

QY 1 GGGGGIEGPTLRQWLAARAGGGGGGG 27
      |||||: | : | : |||||
Db 258 GGGGGIIPG---QSMYMGAGGGGGGAGG 280

RESULT 20
T29167
hypothetical protein T28H11.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29167
R:Neilson, J.; Wohldmann, P.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid T28H11.
A:Reference number: Z20582
A:Accession: T29167
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-405 <NEL>
A:Cross-references: EMBL:U64609; PIDN:AAB04598.1; GSPDB:GN00022; CESP:T28H11.5
A:Experimental source: strain Bristol N2; clone T28H11
C:Genetics:
A:Gene: CESP:T28H11.5
A:Map position: 4

```

RESULT 12
 GCR 17 protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C>Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 24-Sep-1998
 C:Accession: S49194
 R:Parment, C.; Hughes, D.M.; Lloyd, P.; Flavell, A.J.
 submitted to the EMBL Data Library, May 1993
 A:Description: A variety of different glycine repeats in *Drosophila* genes.
 A:Reference number: S49192
 A:Accession: S49194
 A:Molecule type: mRNA
 A:Residues: 1-196 <PAR>
 A:Cross-references: EMBL:X71974; NID:g510501; PID:g510502
 C:Genetics:
 A:Gene: FlyBase:anon-Pen15EF
 A:Cross-references: FlyBase:FBgn0003057
 Query Match 32.1%; Score 72; DB 2; Length 196;
 Best Local Similarity 53.3%; Pred. No. 1.3;
 Matches 16; Conservative 2; Mismatches 10; Indels 2; Gaps 1;
 QY 1 GGGGIEGPTLRQWLAARAGGGGGGGGIEG 30
 DB 57 GLGGGLGG--LSGGGLGKGLGGGGGGGGYSG 84
 RESULT 13
 S16063
 acp-22 protein - yellow mealworm
 C:Species: *Tenebrio molitor* (yellow mealworm)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
 C:Accession: S16063
 R:Bouhin, H.; Charles, J.P.; Quennedey, B.; Delachambre, J.
 submitted to the EMBL Data Library, July 1991
 A:Description: Molecular cloning of a glycine-rich cuticular protein.
 A:Reference number: S16063
 A:Accession: S16063
 A:Molecule type: mRNA
 A:Residues: 1-199 <BOU>
 A:Cross-references: EMBL:X60455; NID:g10750; PID:g10751
 Query Match 32.1%; Score 72; DB 2; Length 199;
 Best Local Similarity 48.6%; Pred. No. 1.4;
 Matches 17; Conservative 1; Mismatches 9; Indels 8; Gaps 1;
 QY 1 GGGGIEGPTLRQW-----LAARAGGGGGGGG 27
 DB 65 GGGGEGEGEGREHRLGGGLELGGGGGGGGGGG 99
 RESULT 14
 S32224
 acp-22 protein - yellow mealworm
 C:Species: *Tenebrio molitor* (yellow mealworm)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
 C:Accession: S32224
 R:Bouhin, H.; Braquet, C.; Charles, J.P.; Delachambre, J.
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S32224
 A:Accession: S32224
 A:Molecule type: DNA
 A:Residues: 1-199 <BOU>
 A:Cross-references: EMBL:X72783; NID:g288439; PID:g288440
 C:Genetics:
 A:Introns: 5/3; 131/3
 Query Match 32.1%; Score 72; DB 2; Length 199;

Best Local Similarity 48.6%; Pred. No. 1.4;
 Matches 17; Conservative 1; Mismatches 9; Indels 8; Gaps 1;
 QY 1 GGGGIEGPTLRQW-----LAARAGGGGGGGG 27
 DB 65 GGGGEGEGEGREHRLGGGLELGGGGGGGGGGG 99
 RESULT 15
 A34466
 calpain (EC 3.4.22.17) light chain - bovine
 C:Species: *Bos primigenius taurus* (cattle)
 C>Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 16-Jul-1999
 C:Accession: A34466
 R:McClelland, P.; Lash, J.A.; Hathaway, D.R.
 J. Biol. Chem. 264, 17428-17431, 1989
 A:Title: Identification of major autolytic cleavage sites in the regulatory subunit of
 A:Reference number: A34466; MUID:90008911
 A:Accession: A34466
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-263 <WCC>
 A:Cross-references: GB:J05065; NID:g162780; PID:AAA30422.1; PID:g162781
 C:Superfamily: calpain small chain; calmodulin repeat homology
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer;
 F:1-51/Domain: glycine-rich <GLY>
 F:91-122/Domain: calmodulin repeat homology <EF1>
 F:134-166/Domain: calmodulin repeat homology <EF2>
 F:167-196/Domain: calmodulin repeat homology <EF3>
 F:199-231/Domain: calmodulin repeat homology <EF4>
 F:232-263/Domain: calmodulin repeat homology <EF5>
 Query Match 32.1%; Score 72; DB 2; Length 263;
 Best Local Similarity 55.2%; Pred. No. 1.7;
 Matches 16; Conservative 3; Mismatches 8; Indels 2; Gaps 1;
 QY 1 GGGGIEG--PTLRQWLAARAGGGGGGGG 27
 DB 15 GGGGGLGGGLGNVLGGGLISGAGGGGGGGG 43
 RESULT 16
 C1PGL
 calpain (EC 3.4.22.17) small chain - pig
 N:Alternate names: calcium-activated neutral proteinase (CANP); calpain light chain;
 C:Species: *Sus scrofa domestica* (domestic pig)
 C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
 C:Accession: A25166; B25166
 R:Sakihama, T.; Kakidani, H.; Zenita, K.; Yumoto, N.; Kikuchi, T.; Sasaki, T.; Kannag
 Proc. Natl. Acad. Sci. U.S.A. 82, 6075-6079, 1985
 A:Title: A putative Ca²⁺-binding protein: structure of the light subunit of porcine c
 A:Reference number: A25166; MUID:85298299
 A:Accession: A25166
 A:Molecule type: mRNA
 A:Residues: 1-266 <SAK>
 A:Cross-references: GB:M11778; NID:g164402; PID:AAA31010.1; PID:g164403; GB:M11779;
 A:Accession: B25166
 A:Molecule type: protein
 A:Residues: 2-56; 125-143; 157-177; 247-248; 250-256; 265-266 <SA2>
 C:Complex: heterodimer of L (large) and S (small) chains
 C:Function:
 A:Description: catalyzes the hydrolysis of peptides
 A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and b
 C:Superfamily: calpain small chain; calmodulin repeat homology
 C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; duplication;
 F:1-54/Domain: glycine-rich <GLY>
 F:94-125/Domain: calmodulin repeat homology <EF1>
 F:137-169/Domain: calmodulin repeat homology <EF2>
 F:170-199/Domain: calmodulin repeat homology <EF3>
 F:202-234/Domain: calmodulin repeat homology <EF4>
 F:235-266/Domain: calmodulin repeat homology <EF5>
 F:1/Modified site: acetylated amino end (Met) #status experimental

Best Local Similarity 59.3%; Pred. No. 0.2;
Matches 16; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 1 GGGGIEGPTLQWLAAARAGGGGGG 27
||||| | : : : |||||
Db 354 GGGGIPQSV--YMGAGGGGGGGG 378

RESULT 3
S20099
transforming protein jund - chicken
C:Species: Gallus gallus (chicken)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S20099
R:Hartl, M.; Hutchins, J.T.; Vogt, P.K.
Oncogene 6, 1623-1631, 1991
A:Title: The chicken jund gene and its product.
A:Reference number: S20099; MUID:92019832
A:Accession: S20099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <R>
A:CROSS-references: EMBL:X60063; NID:g62927; PIDN:CAA42665.1; PID:g62928
C:Superfamily: jun transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; nucleus; transcription regulation
F:237-277/Domain: fos/jun DNA-binding domain homology <FUD>

Query Match 34.8%; Score 78; DB 2; Length 323;
Best Local Similarity 57.6%; Pred. No. 0.53;
Matches 19; Conservative 0; Mismatches 8; Indels 6; Gaps 1;

QY 1 GGGGIEGPTLQWLAAARAGGGGGGPTL 33
||||| | : : : |||||
Db 142 GGGGPNNG-----AAAAGGGGGGGGGGEL 168

RESULT 4
A44805
eggshell protein precursor - fluke (Schistosoma haematobium)
C:Species: Schistosoma haematobium
C:Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 05-May-2000
C:Accession: A44805; C44805
R:Bobek, L.A.; LoVerde, P.T.; Rekosh, D.M.
Exp. Parasitol. 68, 17-30, 1989
A:Title: Schistosoma haematobium: analysis of eggshell protein genes and their expression
A:Reference number: A44805; MUID:89137380
A:Accession: A44805
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-220 <BOB>
A:CROSS-references: GB:M27659
A:Accession: C44805
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-70,96-140,142-220 <BO2>
A:CROSS-references: GB:M27658; NID:g160978; PID:g160979
C:Superfamily: fluke eggshell protein
C:Keywords: egg shell
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-220/Product: eggshell protein #status predicted <MAT>

Query Match 34.4%; Score 77; DB 2; Length 220;
Best Local Similarity 53.3%; Pred. No. 0.47;
Matches 16; Conservative 0; Mismatches 4; Indels 10; Gaps 1;

QY 1 GGGGIEGPTLQWLAAARAGGGGGGGG 30
||||| | : : : |||||
Db 55 GGGGIEG-----GGNGGGGGGGG 74

RESULT 5
T27609
hypothetical protein ZC477.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T27609
R:Du, Z.
Submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid ZC477.
A:Reference number: Z20392
A:Accession: T27609
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-307 <DUZ>
A:CROSS-references: EMBL:U40802; PIDN:AAA81510.1; CESP:ZC477.1
C:Genetics:
A:Gene: CESP:ZC477.1
A:Introns: 32/1; 275/1
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 34.4%; Score 77; DB 2; Length 307;
Best Local Similarity 63.0%; Pred. No. 0.63;
Matches 17; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 1 GGGGIEGPTLQWLAAARAGGGGGG 27
||||| | : : : |||||
Db 275 GGGGIPG----QSVYMGAGGGGGGG 297

RESULT 6
T49109
glycine-rich protein - Arabidopsis thaliana
N:Alternate names: protein AT4g22020
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C:Accession: T49109
R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.
Submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25016
A:Accession: T49109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-396 <BRV>
A:CROSS-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4g22020
A:Experimental source: cultivar Columbia; BAC clone FIN20
C:Genetics:
A:Gene: ATSP:AT4g22020
A:Map position: 4
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 33.7%; Score 75.5; DB 2; Length 396;
Best Local Similarity 53.3%; Pred. No. 1.1;
Matches 16; Conservative 1; Mismatches 6; Indels 7; Gaps 1;

QY 1 GGGGIEGPTLQWLAAARAGGGGGG 30
||||| | : : : |||||
Db 239 GAGGGVSG-----AAGGGGGGGGGG 261

RESULT 7
JC6087
helix-loop-helix transcription factor, BETA3 - hamster
C:Species: Cricetinae gen. sp. (hamster)
C:Date: 28-May-1997 #sequence_revision 18-Jul-1997 #text_change 05-Nov-1999
C:Accession: JC6087
R:Peyton, M.; Stellrecht, C.M.W.; Naya, F.J.; Huang, H.P.; Samora, P.J.; Tsai, M.J.
Mol. Cell. Biol. 16, 626-633, 1996
A:Title: BETA3, a novel helix-loop-helix protein, can act as a negative regulator of
A:Reference number: JC6087; MUID:96140430
A:Accession: JC6087
A:Molecule type: mRNA
A:Residues: 1-367 <PEY>
A:CROSS-references: GB:S80870; NID:g1911496; PIDN:AAB50691.1; PID:g1911497
A:Experimental source: insulin tumor cell
C:Comment: This factor belongs to the tissue-specific class B basic helix-loop-helix
ssion.

Wed Oct 9 10:30:19 2002

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 9, 2002, 08:54:17 ; Search time 9.2178 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838c-34
Perfect score: 224
Sequence: 1 GGGGGIEGPTLRQWLAARAG.....GGGGGIEGPTLRQWLAARA 41

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	37.9	500	2 T20961	hypothetical prote
2	83	37.1	388	2 T29173	hypothetical prote
3	78	34.8	323	2 S20099	transforming prote
4	77	34.4	220	2 A44805	eggshell protein p
5	77	34.4	307	2 T27609	hypothetical prote
6	75.5	33.7	396	2 T49109	glycine-rich prote
7	75	33.5	367	2 T6087	helix-loop-helix t
8	74.5	33.3	346	1 S35500	heterogeneous ribo
9	74	33.0	228	2 D86416	probable beta-1,3
10	74	33.0	239	2 S49193	GCR 101 protein -
11	73	32.6	270	2 G84728	hypothetical prote
12	72	32.1	196	2 S49194	GCR 17 protein -
13	72	32.1	199	2 S16063	acp-22 protein - y
14	72	32.1	199	2 S32224	acp-22 protein - y
15	72	32.1	263	2 A34466	calpain (EC 3.4.22
16	72	32.1	266	1 C1RBL	calpain (EC 3.4.22
17	72	32.1	266	1 C1HUL	calpain (EC 3.4.22
18	72	32.1	268	1 C1HUL	calpain (EC 3.4.22
19	72	32.1	290	2 T23416	hypothetical prote
20	72	32.1	405	2 T29167	hypothetical prote
21	72	32.1	433	2 S20963	homeotic protein H
22	72	32.1	1028	2 A56038	DNA-binding protei
23	72	32.1	1213	2 S16356	ovo protein - frui
24	71.5	31.9	201	2 T49792	hypothetical prote
25	71.5	31.9	404	2 S54729	RNA-binding protei
26	71.5	31.9	1473	2 T13855	suppressor of sabl
27	71.5	31.9	1585	2 T31611	hypothetical prote
28	71	31.7	371	2 T13021	hypothetical prote
29	71	31.7	490	2 T09084	phosphatidylinosit

30	71	31.7	496	2 H70839	hypothetical glyci
31	71	31.7	534	2 JC4572	signal recognition
32	71	31.7	767	2 E70895	hypothetical glyci
33	71	31.7	892	2 T27005	hypothetical prote
34	71	31.7	1325	2 T13386	hypothetical prote
35	70.5	31.5	431	1 WJHU2G	homeotic protein H
36	70	31.2	255	2 B84777	hypothetical prote
37	70	31.2	386	1 S22315	snRNP-associated p
38	70	31.2	1660	2 A70869	hypothetical glyci
39	70	31.2	2174	2 E5965	hypothetical glyci
40	69.5	31.0	369	1 TTFVAF	transforming prote
41	69	30.8	201	2 T00799	hypothetical prote
42	69	30.8	221	2 T04592	glycine-rich cell
43	69	30.8	291	1 S31415	glycine-rich prote
44	69	30.8	309	2 T19389	hypothetical prote
45	69	30.8	342	2 S14432	heterogeneous ribo

ALIGNMENTS

RESULT 1

T20961
hypothetical protein FL5B9.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20961
R:Percy, C.
Submitted to the EMBL Data Library, August 1996
A:Reference number: Z19351
A:Accession: T20961
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-500 <WIL>
A:Cross-references: EMBL:Z78013; PIDN:CA01420.1; GSPDB:GN00023; CESP:FL5B9.5
A:Experimental source: clone FL5B9
C:Genetics:
A:Gene: CESP:FL5B9.5
A:Map position: 5
A:Introns: 46/3; 63/3; 125/2; 283/3; 391/1; 446/1

Query Match 37.9%; Score 85; DB 2; Length 500;
Best Local Similarity 55.2%; Pred. No. 0.16;
Matches 16; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 2 GGGGIEGPTLRQWLAARAGGGGGGIEG 30
||| | : : : | : : : : | : : : : |
Db 423 GGGAAAGSMIGRFLSNRGGGGGGG 451

RESULT 2

T29173
hypothetical protein T28H11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29173
R:Nelson, J.; Wohldmann, P.
Submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid T28H11.
A:Reference number: Z20582
A:Accession: T29173
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-388 <NEL>
A:Cross-references: EMBL:U64609; PIDN:AB04604.1; GSPDB:GN00022; CESP:T28H11.1
A:Experimental source: strain Bristol N2; clone T28H11
C:Genetics:
A:Gene: CESP:T28H11.1
A:Map position: 4
A:Introns: 354/1
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8
Query Match 37.1%; Score 83; DB 2; Length 388;

Wed Oct 9 10:30:19 2002

us-09-422-838c-34.ra1

Page 14

Job time : 6.81733 secs

• • • • •

us-09-422-838c-34.ra1

Wed Oct 9 10:30:19 2002

Query Match 32.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLRQWLAARA 19
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQWLAARA 15

RESULT 30
US-09-516-704-18
; Sequence 18, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
; US-09-516-704-18

Query Match 32.6%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLRQWLAARA 19
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQWLAARA 14

Search completed: October 9, 2002, 09:06:34

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 32.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLRQWLAARA 19
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQWLAARA 15

RESULT 29
US-09-244-298A-232
; Sequence 232, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; Yim, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-232

```

; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
;
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
;
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 220:
;
; US-08-973-225-220

```

```

Query Match 32.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 1EGPTLROWLAARA 19
| | | | | | | | | | | | | | | |
DB 2 1EGPTLROWLAARA 15

```

```

RESULT 27
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Poddaturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
;
; US-09-244-298A-18

```

```

Query Match 32.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 1EGPTLROWLAARA 19
| | | | | | | | | | | | | | | |
DB 1 1EGPTLROWLAARA 14

```

```

RESULT 28
US-09-244-298A-19
; Sequence 19, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Poddaturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:

```

us-09-422-838c-34.ra1

Wed Oct 9 10:30:19 2002

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-232

Query Match 32.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014; Indels 0;
Matches 14; Conservative 0; Mismatches 0; Gaps 0;

QY 6 IEGPTLROWLAARA 19
Db 2 IEGPTLROWLAARA 15

RESULT 24
US-08-973-225-18
Sequence 18, Application US/08973225A
Patent No. 6083913

GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haseiden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Beta-ala"

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-08-973-225-18

Query Match 32.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014; Indels 0;
Matches 14; Conservative 0; Mismatches 0; Gaps 0;

QY 6 IEGPTLROWLAARA 19

Db 1 IEGPTLROWLAARA 14

RESULT 25

US-08-973-225-194
Sequence 194, Application US/08973225A

Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwiria, Steven E.

Duffin, David J.

Gates, Christian

Haseiden, Sherril S.

Matheakis, Larry C.

Schatz, Peter J.

Wagstrom, Christopher R.

Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-08-973-225-194

Query Match 32.6%; Score 73; DB 3; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.014;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEGPTLROWLAARA 19

Db 2 IEGPTLROWLAARA 15

RESULT 26

US-08-973-225-220

Sequence 220, Application US/08973225A

Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwiria, Steven E.

Duffin, David J.

Gates, Christian

Haseiden, Sherril S.

CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
US-08-764-640-18

Query Match 32.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLRQWLAARA 19
|||||
Db 1 IEPTLRQWLAARA 14

RESULT 22
US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depirnce, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids

CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
US-08-764-640-194

Query Match 32.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLRQWLAARA 19
|||||
Db 2 IEPTLRQWLAARA 15

RESULT 23
US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depirnce, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185

Query Match 32.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEGPTLQWLAA 19
Db 2 IEGPTLQWLAA 15

RESULT 17
US-09-244-298A-17
; Sequence 17, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-185

TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17

Query Match 32.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEGPTLQWLAA 19
Db 1 IEGPTLQWLAA 14

RESULT 18
US-09-244-298A-185
; Sequence 185, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-185

Query Match 32.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/516,704
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13

Query Match 32.6%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLQWLAA 19
| | | | | | | | | |
Db 1 IEPTLQWLAA 14

RESULT 12

US-09-516-704-193
Sequence 193, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids

TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match 32.6%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLQWLAA 19
| | | | | | | | | |
Db 1 IEPTLQWLAA 14

RESULT 13

US-08-764-640-17
Sequence 17, Application US/08764640
Patent No. 5869451
GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 32.6%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTLQWLAA 19
| | | | | | | | | |

```
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depvince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-13

Query Match 32.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEGPTLRQWLAAARA 19
Db 1 IEGPTLRQWLAAARA 14

RESULT 10
US-09-244-298A-193
; Sequence 193, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depvince, Randolph B.
; APPLICANT: Podduturi, Surekha

US-09-244-298A-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depvince, Randolph B.
; APPLICANT: Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-193

Query Match 32.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEGPTLRQWLAAARA 19
Db 1 IEGPTLRQWLAAARA 14

RESULT 11
US-09-516-704-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depvince, Randolph B.
; APPLICANT: Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193

Query Match 32.6%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTTLQWLAA 19
DB 1 IEPTTLQWLAA 14

RESULT 7
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>

Query Match 32.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTTLQWLAA 19
DB 1 IEPTTLQWLAA 14

RESULT 9

TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 32.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IEPTTLQWLAA 19
DB 1 IEPTTLQWLAA 14

RESULT 8
US-08-973-225-193
Sequence 193, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-08-973-225-193

Query Match 32.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
TITLE OF INVENTION: TO SAID POLYPEPTIDE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: 812-5 Hirano
STREET: Isshinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 33.3%; Score 74.5; DB 3; Length 738;
Best Local Similarity 55.0%; Pred. No. 0.48; Mismatches 15; Indels 3; Gaps 2;
Matches 22; Conservative 0;

Qy 1 GGGGGIEGPTLQWLAARAGGG-GGGGGIEGPTLQWLA 39
||||| | ||||| ||||| | |||||
Db 450 GGGGGAGALAAALAAAGAGGGLGGGG--GGALAAALAA 487

RESULT 5
US-08-764-640-13
Sequence 13, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 32.6%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 IEGPTLQWLAARA 19
||||| ||||| |||||
Db 1 IEGPTLQWLAARA 14

RESULT 6
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /product= "Ava"
US-08-764-640-231

Query Match 34.2%; Score 76.5; DB 2; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0091;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 7 EGPTLROWLAARAGGGGGGEGPTLROWLA 38
:||||:|:|:| | :||||:|:|:| |
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 2

US-09-244-298A-231
; Sequence 231, Application US/09244298A
; Patent No. 6121238

; GENERAL INFORMATION:
; APPLICANT: Dower, William J.

; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

; NUMBER OF SEQUENCES: 244

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398

; CITY: Research Triangle Park
; STATE: NC

; COUNTRY: USA
; ZIP: 27709

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A

; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.

; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000

; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 amino acids
; TYPE: amino acid

; STRANDEDNESS:
; TOPOLOGY: linear

; MOLECULE TYPE: peptide
; FEATURE:

; NAME/KEY: Modified-site
; LOCATION: 13

; OTHER INFORMATION: /product= "Ava"
US-09-244-298A-231

Query Match 34.2%; Score 76.5; DB 3; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0091;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 7 EGPTLROWLAARAGGGGGGEGPTLROWLA 38

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24
:||||:|:|:| | :||||:|:|:| |

RESULT 3

US-09-516-704-231
; Sequence 231, Application US/09516704
; Patent No. 6251864

; GENERAL INFORMATION:
; APPLICANT: Dower, William J.

; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

; NUMBER OF SEQUENCES: 244

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398

; CITY: Research Triangle Park
; STATE: NC

; COUNTRY: USA
; ZIP: 27709

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704

; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.

; REGISTRATION/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000

; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 amino acids
; TYPE: amino acid

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear

; MOLECULE TYPE: peptide
; FEATURE:

; NAME/KEY: Modified-site
; LOCATION: 13

; OTHER INFORMATION: /product= "Ava"
; SEQUENCE DESCRIPTION: SEQ ID NO: 231:
US-09-516-704-231

Query Match 34.2%; Score 76.5; DB 4; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0091;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 7 EGPTLROWLAARAGGGGGGEGPTLROWLA 38
:||||:|:|:| | :||||:|:|:| |

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 4

US-08-864-038A-3

; Sequence 3, Application US/08864038A
; Patent No. 6001592

; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 9, 2002, 08:55:27 ; Search time 6.81733 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838C-34
Perfect score: 224
Sequence: 1 GGGGGIEGPTLRQWLAARAG.....GGGGGIEGPTLRQWLAARA 41

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	76.5	34.2	25	2	US-08-764-640-231
2	76.5	34.2	25	3	US-09-244-298A-231
3	76.5	34.2	25	4	US-09-516-704-231
4	74.5	33.3	738	3	US-08-864-038A-3
5	73	32.6	14	2	US-08-764-640-193
6	73	32.6	14	2	US-08-764-640-193
7	73	32.6	14	3	US-08-973-225-193
8	73	32.6	14	3	US-08-973-225-193
9	73	32.6	14	3	US-09-244-298A-193
10	73	32.6	14	4	US-09-516-704-193
11	73	32.6	14	4	US-09-516-704-193
12	73	32.6	15	2	US-08-764-640-17
13	73	32.6	15	2	US-08-764-640-17
14	73	32.6	15	3	US-08-764-640-185
15	73	32.6	15	3	US-08-973-225-17
16	73	32.6	15	3	US-08-973-225-185
17	73	32.6	15	3	US-09-244-298A-17
18	73	32.6	15	3	US-09-244-298A-185
19	73	32.6	15	4	US-09-516-704-17
20	73	32.6	15	4	US-09-516-704-185
21	73	32.6	16	2	US-08-764-640-18
22	73	32.6	16	2	US-08-764-640-194
23	73	32.6	16	2	US-08-764-640-232
24	73	32.6	16	3	US-08-973-225-18
25	73	32.6	16	3	US-08-973-225-194
26	73	32.6	16	3	US-08-973-225-220
27	73	32.6	16	3	US-09-244-298A-18

28	73	32.6	16	3	US-09-244-298A-194
29	73	32.6	16	3	US-09-244-298A-232
30	73	32.6	16	4	US-09-516-704-18
31	73	32.6	16	4	US-09-516-704-194
32	73	32.6	16	4	US-09-516-704-232
33	72.5	32.4	552	2	US-08-317-401E-4
34	72	32.1	266	4	US-09-032-523-7
35	72	32.1	268	2	US-08-835-099A-9
36	72	32.1	268	3	US-09-157-349-9
37	71	31.7	534	2	US-08-317-401E-2
38	71	31.7	870	2	US-09-010-328B-2
39	70.5	31.5	201	4	US-09-052-995-1
40	70.5	31.5	201	4	US-09-053-003-40
41	69	30.8	14	2	US-08-764-640-195
42	69	30.8	14	2	US-08-764-640-199
43	69	30.8	14	3	US-08-973-225-195
44	69	30.8	14	3	US-08-973-225-199
45	69	30.8	14	3	US-09-244-298A-195

ALIGNMENTS

RESULT 1
US-08-764-640-231
Sequence 231, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:

CC to AAA69526 and AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX

SQ Sequence 35 AA;
Query Match 79.28; Score 177.5; DB 21; Length 35;
Best Local Similarity 97.28; Pred. No. 5.2e-14;
Matches 35; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 IEGPTLQWLAAAGGGGGGIEGPTLQWLAAARA 41
|||||
Db 1 IEGPTLQWLAAARA-GGGGGGIEGPTLQWLAAARA 35

RESULT 26
AAB17302
ID AAB17302 standard; Peptide; 40 AA.
XX AC AAB17302;

XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:358.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

XX Example 1; Page 322; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 40 AA;

Query Match 77.78; Score 174; DB 21; Length 40;
Best Local Similarity 87.58; Pred. No. 1.5e-13;
Matches 35; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY 6 IEGPTLQWLAAAGGGGGGIEGPTLQWLAAARA 41
|||||
Db 1 IEGPTLQWLAAAGGGGGGIEGPTLQWLAAARA 40

RESULT 27
AAB17291
ID AAB17291 standard; Peptide; 34 AA.

XX AC AAB17291;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:347.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

XX Example 1; Page 317; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 34 AA;

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 42 AA;
 SQ Query Match 80.8%; Score 181; DB 21; Length 42;
 Best Local Similarity 85.7%; Pred. No. 2.4e-14;
 Matches 36; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 6 IEGPTLRQWLAARA-----GGGGGGGIEGPTLRQWLAARA 41
 |||||
 Db 1 IEGPTLRQWLAARAGGGGGGGGIEGPTLRQWLAARA 42

RESULT 25
 AAB17292
 ID AAB17292 standard; Peptide; 35 AA.

XX AAB17292;

AC AAB17292;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:348.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 317-318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TRPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;
 SQ Query Match 81.2%; Score 182; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 1.6e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 41
 |||||
 Db 1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 24
 AAB17296
 ID AAB17296 standard; Peptide; 42 AA.

XX AAB17296;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:352.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX

SQ Sequence 39 AA;

Query Match 81.5%; Score 182.5; DB 21; Length 39;
 Best Local Similarity 92.3%; Pred. No. 1.5e-14;
 Matches 36; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 6 IEGPTLQWLAAARAGG---GGGGGIEGPTLQWLAAARA 41
 |||||
 Db 1 IEGPTLQWLAAARAGGCPGGGGGIEGPTLQWLAAARA 39

RESULT 22

AAB17306
 ID AAB17306 standard; Peptide: 36 AA.

XX
 AC AAB17306;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:362.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

PN 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 324; 608pp; English.

PS The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-Fi-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX

SQ Sequence 36 AA;

Query Match 81.2%; Score 182; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 1.6e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 IEGPTLQWLAAARAGGGGGGIEGPTLQWLAAARA 41
 |||||
 Db 1 IEGPTLQWLAAARAGGGGGGIEGPTLQWLAAARA 36

RESULT 23

AAY96526
 ID AAY96526 standard; peptide: 36 AA.

XX
 AC AAY96526;

DT 04-SEP-2000 (first entry)

DE Thrombopoietin mimetic peptide compound 7.

XX Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

XX Key Modified-site 1 Location/Qualifiers
 FT /note= "optionally linked to an Fc molecule"

FT Peptide 1..14

FT Peptide /label= TMP_1

FT Peptide 15..18

FT Peptide /label= linker

FT Peptide 19..32

XX /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-nTMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
 CC X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
 CC X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
 CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
 CC or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
 CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
 CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

CC	0 or 1, provided that at least 1 of a and b is 1. The composition can
CC	have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC	activities. DNAs, vectors and host cells from the present invention can
CC	be used for producing pharmaceutical compositions. The compositions are
CC	useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC	The use of an Fc domain (rather than a Fab domain) can provide a longer
CC	half-life or incorporate functions such as Fc receptor binding, protein
CC	A binding, complement fixation, and possibly placental transfer. AAA69443
CC	to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC	sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 38 AA;
	Query Match 81.7%; Score 183; DB 21; Length 38;
	Best Local Similarity 94.7%; Pred. No. 1.3e-14;
	Matches 36; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY	6 IEGPTLRQLAARA--GGGGGGGIEGPTLRQLAARA 41
Db	1 IEGPTLRQLAARAGGGGGGGGIEGPTLRQLAARA 38
RESULT 20	
AAB17304	ID AAB17304 standard; Peptide; 39 AA.
XX	AAB17304;
XX	31-OCT-2000 (first entry)
DE	TPO-mimetic peptide sequence SEQ ID NO:360.
XX	Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW	autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW	immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW	MMP; inhibitor; erythropoietin; thrombospondin; interleukin 1;
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW	vascular endothelial growth factor; matrix metalloproteinase;
KW	asthma; thrombosis; pharmaceutical.
XX	Synthetic.
OS	
XX	WO200024782-A2.
PN	04-MAY-2000.
XX	
PD	25-OCT-1999; 99WO-US25044.
XX	
PF	23-OCT-1998; 98US-0105371.
XX	22-OCT-1999; 99US-0428082.
PR	(AMGE-) AMGEN INC.
PR	
XX	
PA	
XX	
PI	Feige U, Liu C, Cheetham J, Boone TC;
XX	
DR	WPI; 2000-350702/30.
XX	
PT	Novel composition of matter comprising an Fc domain and
PT	pharmacologically active peptides, useful for treating cancer and
PT	autoimmune diseases -
XX	
P5	Example 1; Page 323; 608pp; English.
XX	
CC	The present invention describes composition of matter (I) comprising an
CC	Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC	(X1)-a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC	independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC	-(L1)c-P1-(L2)d-P2-(L3)e-p*3, or -(L1)c-P1-(L2)d-P2-(L3)e-p3-(L4)f-p4
CC	where P1, P2, P3, and P4 = are each independently sequences of
CC	pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC	independently linkers; and a, b, c, d, e, and f = are each independently
CC	0 or 1, provided that at least 1 of a and b is 1. The composition can
CC	have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC 10 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂,
 CC X₂-X₁-L₃, X₂-X₁-L₄, X₁-X₁-L₀, X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and
 CC X₁-X₁-L₄. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K,
 CC T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TmPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.,
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 82.6%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 7.2e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 41
 |||||
 Db 1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 18

AAB17294

ID AAB17294 standard; Peptide; 37 AA.

XX AC AAB17294;

XX DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:350.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X₁)a-Fi-(X₂)b, where: F1 = an Fc domain; X₁ and X₂ = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. Amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 37 AA;

Query Match 81.9%; Score 183.5; DB 21; Length 37;
 Best Local Similarity 97.3%; Pred. No. 1.1e-14;
 Matches 36; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 IEGPTLRQWLAARA-GGGGGGIEGPTLRQWLAARA 41
 |||||
 Db 1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 37

RESULT 19

AAB17295

ID AAB17295 standard; Peptide; 38 AA.

XX AC AAB17295;

XX DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:351.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X₁)a-Fi-(X₂)b, where: F1 = an Fc domain; X₁ and X₂ = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;
 SQ Query Match 82.6%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 7.2e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 6 IEGPTLQWLAAARAGGGGGGIEGPTLQWLAAARA 41
 Db 1 IEGPTLQWLAAARAGGGGGGIEGPTLQWLAAARA 36
 |||||||||||||||||||||||||||||||||||||||||

RESULT 16
 AAB17307
 ID AAB17307 standard; Peptide; 36 AA.
 XX AC
 XX AAB17307;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:363.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 324; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

SQ Query Match 82.6%; Score 185; DB 21; Length 36;

Best Local Similarity 97.2%; Pred. No. 7.2e-15;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 IEGPTLQWLAAARAGGGGGGIEGPTLQWLAAARA 41

Db 1 IEGPTLQWLAAARAGGGGGGIEGPTLQWLAAARA 36

|||||||||||||||||||||||||||||||||||||||

RESULT 17

AA996524

ID AA996524 standard; peptide; 36 AA.

XX AC

XX AA996524;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 5.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX Synthetic.

XX Key Location/Qualifiers

XX Modified-site 1 /note= "optionally linked to an Fc molecule"

XX Peptide 1..14 /label= TMP_1

XX Disulfide-bond 9..31 /note= "optional"

XX Peptide 15..22 /label= linker

XX Peptide 23..36 /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP_1-(L_1)-TMP_2),
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antitumoric, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

```

SQ      Sequence      36 AA;

Query Match      83.0%; Score 186; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 5.5e-15;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY      6 IEGPTLROWLAARAGGGGGGGGIEGPTLROWLAARA 41
Db      1 IEGPTLROWLAARAGGGGGGGGIEGPTLROWLAARA 36

```

RESULT 14	
AAY96523	
ID	AAY96523 standard; peptide; 36 AA.
XX	
XX	AAY96523;
XX	
XX	04-SEP-2000 (first entry)
DT	
XX	
XX	Thrombopoietin mimetic peptide compound 4.
DE	
XX	
XX	Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW	anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW	immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX	
OS	Synthetic.
XX	
Key	Location/Qualifiers
FT	Modified-site 1
FT	/note= "optionally linked to an Fc molecule"
FT	1..14
FT	/label= TMP_1
FT	15..22
FT	/label= linker
FT	
FT	Modified-site 18
FT	/note= "optionally modified by bromoacetyl or PEG"
FT	23..36
FT	/label= TMP_2
FT	

10 to 14 residues in length comprising X₂-X₁-0, X₂-X₁-1, X₂-X₁-2, X₂-X₁-3, X₂-X₁-4, X₁-X₁-0, X₁-X₁-1, X₁-X₁-2, X₁-X₁-3, and X₁-X₁-4. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N, or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K, T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

Query Match	83.0%	Score 186;	DB 21;	Length 36;
Best Local Similarity	97.2%	Pred. No. 5.5e-15;		
Matches	35:	Conservative	0: Mismatches	1: Indels
			0: Gaps	0: Gaps

Qy 6 IEGPTLROWLAARAGGGGGGIEGPTLROWLAARA 41
 |||
 Db 1 IEGPTLROWLAARAGGGKGGGIEGPTLROWLAARA 36

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 60 AA;
Query Match 86.6%; Score 194; DB 21; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41
|||||
DB 2 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37
|||||
RESULT 12
AAB16960
ID AAB16960 standard; Protein; 269 AA.
XX
AC AAB16960;
XX
DT 31-OCT-2000 (first entry)
XX
DE TMP-TMP-Fc protein sequence SEQ ID NO:10.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
DR N-PSDB; AAA69446.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 2; Page 185-186; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 60 AA;
Query Match 86.6%; Score 194; DB 21; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41
|||||
DB 2 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37
|||||
RESULT 13
AAB17301
ID AAB17301 standard; Peptide; 36 AA.
XX
AC AAB17301;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:357.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
DR N-PSDB; AAA69446.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 321; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently

where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently active linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antitasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences of the present invention.

[illegible]

```

RESULT 8
AAB17293
ID AAB17293 standard; Peptide; 36 AA.
XX
XX
AC
AAB17293;
XX
XX
DT
31-OCT-2000 (first entry)
XX
DE
TPO-mimetic peptide sequence SEQ ID NO:349.
XX
XX

```

KW	Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW	autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW	immunopressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW	MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW	vascular endothelial growth factor; matrix metalloproteinase;
KW	asthma; thrombosis; pharmaceutical.
XX	
XX	Synthetic.
XX	
PN	WO200024782-A2.
XX	
XX	04 - MAY - 2000.
PD	
XX	

PF	25-OCT-1999;	99WO-US25044 .
XX		
PR	23-OCT-1998;	98US-0105371 .
PR	22-OCT-1999;	99US-0428082 .
XX		
XX	(AMGE-) AMGEN INC.	
PA		
XX	Feige U, Liu C, Cheetham J, Boone TC;	
PI		
XX		
DR	WPI; 2000-350702/30 .	
XX		
XX	Novel composition of matter comprising an Fc domain and	
PT	pharmacologically active peptides, useful for treating cancer and	
PT	autoimmune diseases .	

xx The present invention describes composition of matter (I) comprising an
ps CC domain, pharmacologically active peptides, and linkers. Where (I) is:
xx CC F1-a-F1-(X2)^b, where: F1 = an FC domain; X1 and X2 = are each
CC independently selected from -(L1)c-p1, -(L1)c-p1-(L2)d-p2,
CC -(L1)c-p1-(L2)d-p2-(L3)e-p3, or -(L1)c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4
CC where p1, p2, p3, and p4 = are each independently sequences of
CC -(L1)c-p1, -(L1)c-p1-(L2)d-p2, -(L1)c-p1-(L2)d-p2-(L3)e-p3, and L4 = are each

PS WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases

XX

PS Example 2; Page 327; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-E1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; and a, b, c, d, e, and f = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 100.0%; Score 224; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 2.6e-19;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41

Db 2 GGGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 5

AA96531

ID AA96531 standard; Protein; 269 AA.

AC AA96531;

DT 04-SEP-2000 (first entry)

DE Human IgG1 Fc TMP fusion protein.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;

KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;

KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX Homo sapiens.

OS

XX WO200024770-A2.

PN

XX 04-MAY-2000.

PD

XX 22-OCT-1999; 99WO-US24834.

PF

XX 23-OCT-1998; 98US-0105348.

PR

XX (AMGE-) AMGEN INC.

PA

XX Liu C, Feige U, Cheetham J;

PI

XX WPI; 2000-365108/31.

DR

XX N-PSDB; AAA29229.

DR

XX Thrombopoietic peptides which activate mpl receptors and increase the

PT production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

XX

PS Example 2A; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin

XX mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_TMP_2],

CC

DR WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases

XX

PS Example 2; Page 327; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-E1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; and a, b, c, d, e, and f = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 100.0%; Score 224; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 2.6e-19;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41

Db 2 GGGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 4

AA96530

ID AA96530 standard; Protein; 42 AA.

AC AA96530;

DT 04-SEP-2000 (first entry)

DE Thrombopoietin mimetic peptide.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;

KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;

KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX Synthetic.

OS

XX WO200024770-A2.

PN

XX 04-MAY-2000.

PD

XX 22-OCT-1999; 99WO-US24834.

PF

XX 23-OCT-1998; 98US-0105348.

PR

XX (AMGE-) AMGEN INC.

PA

XX Liu C, Feige U, Cheetham J;

PI

XX WPI; 2000-365108/31.

DR

XX N-PSDB; AAA29225.

DR

XX Thrombopoietic peptides which activate mpl receptors and increase the

PT production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

XX

(AMGE-) AMGEN INC.
 Liu C, Feige U, Cheetham J;
 WPI; 2000-365108/31.
 Thrombopoietic peptides which activate mpl receptors and increase the production of platelets or platelet precursors, useful for treatment of diseases which involve thrombocytopenia
 Claim 16; Page 65; 91pp; English.
 A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP₁-(L₁L₁)_TMP₂], is new. TMP₁ and TMP₂ are amino acid sequences varying from at least 10 to 14 residues in length comprising X₂-X₁-L₁-L₁, X₂-X₁-L₁-L₂, X₂-X₁-L₁-L₃, X₂-X₁-L₁-L₄, X₁-X₁-L₁-L₁, X₁-X₁-L₁-L₂, X₁-X₁-L₁-L₃, and X₁-X₁-L₁-L₄. X₁ = I, A, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N, or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K, T, V, N, O or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases (e.g. thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus).

Query Match 100.0%; Score 224; DB 21; Length 41;
 Best Local Similarity 100.0%; Pred. No. 2.5e-19;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41
 |||||
 Db 1 GGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41
 |||||

RESULT 2
 AAB17281
 ID AAB17281 standard; Peptide: 42 AA.
 AC AAB17281;
 XX 31-OCT-2000 (first entry)
 DE TPO-mimetic peptide sequence SEQ ID NO:337.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200024782-A2.
 PN 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 PF 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 PA Feige U, Liu C, Cheetham J, Boone TC;
 PI

XX WPI; 2000-350702/30.
 DR Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases.
 PT
 XX Disclosure; Page 313; 608pp; English.
 XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAG69443
 CC to AAG69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX Sequence 42 AA;
 SQ

Query Match 100.0%; Score 224; DB 21; Length 42;
 Best Local Similarity 100.0%; Pred. No. 2.6e-19;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 41
 |||||
 Db 2 GGGGIEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42
 |||||

RESULT 3
 AAB17308
 ID AAB17308 standard; Peptide: 42 AA.
 XX AAB17308;
 AC AAB17308;
 XX 31-OCT-2000 (first entry)
 DE Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200024782-A2.
 PN 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 PF 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 PA Feige U, Liu C, Cheetham J, Boone TC;
 PI

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 18.4356 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-34

Perfect score: 224

Sequence: 1 GGGGGIEGPTLRQWLAARAG.....GGGGGIEGPTLRQWLAARA 41

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224	100.0	41	21	Thrombopoietin mim
2	224	100.0	42	21	TPO-mimetic peptid
3	224	100.0	42	21	Synthetic TMP-TMP
4	224	100.0	42	21	Thrombopoietin mim
5	224	100.0	269	21	Human IgG1 Fc TMP
6	220	98.2	268	21	Thrombopoietin mim
7	194	86.6	36	21	FC-TMP protein
8	194	86.6	36	21	TPO-mimetic peptid
9	194	86.6	36	21	TPO-mimetic peptid
10	194	86.6	42	21	Thrombopoietin mim
11	194	86.6	60	21	TPO-mimetic peptid
					Synthetic TMP-TMP

12	194	86.6	269	21	AA16960	TMP-TMP-Fc protein
13	186	83.0	36	21	AA16960	TPO-mimetic peptid
14	186	83.0	36	21	AA16960	Thrombopoietin mim
15	185	82.6	36	21	AA16960	TPO-mimetic peptid
16	185	82.6	36	21	AA16960	TPO-mimetic peptid
17	185	82.6	36	21	AA16960	Thrombopoietin mim
18	183.5	81.9	37	21	AA16960	TPO-mimetic peptid
19	183	81.7	38	21	AA16960	TPO-mimetic peptid
20	182.5	81.5	39	21	AA16960	TPO-mimetic peptid
21	182.5	81.5	39	21	AA16960	TPO-mimetic peptid
22	182	81.2	36	21	AA16960	TPO-mimetic peptid
23	182	81.2	36	21	AA16960	Thrombopoietin mim
24	181	80.8	42	21	AA16960	TPO-mimetic peptid
25	177.5	79.2	35	21	AA16960	TPO-mimetic peptid
26	174	77.7	40	21	AA16960	TPO-mimetic peptid
27	171	76.3	34	21	AA16960	TPO-mimetic peptid
28	168	75.0	36	21	AA16960	TPO-mimetic peptid
29	168	75.0	36	21	AA16960	TPO-mimetic peptid
30	168	75.0	36	21	AA16960	Cyclic or linear t
31	166	74.1	36	21	AA16960	Linear thrombopoie
32	166	74.1	36	21	AA16960	TPO-mimetic peptid
33	164.5	73.4	33	21	AA16960	TPO-mimetic peptid
34	158	70.5	32	21	AA16960	Thrombopoietin mim
35	156	69.6	34	21	AA16960	TPO-mimetic peptid
36	151.5	67.6	31	21	AA16960	Thrombopoietin mim
37	145	64.7	30	21	AA16960	TPO-mimetic peptid
38	144	64.3	32	21	AA16960	TPO-mimetic peptid
39	144	64.3	32	21	AA16960	Thrombopoietin mim
40	138.5	61.8	29	21	AA16960	TPO-mimetic peptid
41	132	58.9	28	21	AA16960	TPO-mimetic peptid
42	131.5	58.7	29	21	AA16960	TPO-mimetic peptid
43	129.5	57.8	31	21	AA16960	TPO-mimetic peptid
44	129.5	57.8	31	21	AA16960	TPO-mimetic peptid
45	125.5	56.0	29	21	AA16960	TPO-mimetic peptid

ALIGNMENTS

RESULT 1

AA1696528
ID AA1696528 standard; peptide; 41 AA.

AC AA1696528;

DT 04-SEP-2000 (first entry)

DE Thrombopoietin mimetic peptide compound 9.

FW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.

OS Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1
FT Peptide /note= "optionally linked to an Fc molecule"
FT Peptide 6..19
FT Peptide /label= TMP_1
FT Peptide 20..27
FT Peptide /label= linker
FT Peptide 28..41
FT Peptide /label= TMP_2

WO2000024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX


```

Query Match      29.2%; Score 56; DB 11; Length 4833;
Best Local Similarity 50.0%; Pred. No. 9e+02;
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 IEQPTLRQWLAAAGGNGSGGIEGP 26
    :||| : ||| ||| ||| :
Db 7 LEQEGLPGLAAAGAGGSGSALHP 32

RESULT 26
Q9QYX7 PRELIMINARY; PRT; 5038 AA.
AC Q9QYX7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ACZONIN.
GN ACZ OR ACZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilmann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin.";
RN J. Cell Biol. 147:151-162(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Killmann M.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y19185; CAB60731.2; -.
DR HSSP: P04410; 1A25.
DR MGD: MGI:1349390; Acz.
DR InterPro: IPR000008; C2.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PDZ; 1.
DR SMART: SM00239; C2; 2.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR PROSITE: PS50004; C2_DOMAIN_2; 2.
DR PROSITE: PS50106; PDZ; 1.
SQ SEQUENCE 5038 AA; 547600 MW; DADA460CF3B40888 CRC64;

Query Match      29.2%; Score 56; DB 11; Length 5038;
Best Local Similarity 50.0%; Pred. No. 9.4e+02;
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 IEQPTLRQWLAAAGGNGSGGIEGP 26
    :||| : ||| ||| ||| :
Db 7 LEQEGLPGLAAAGAGGSGSALHP 32

RESULT 27
Q9FT25 PRELIMINARY; PRT; 202 AA.
AC Q9FT25;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE P0436E04.1 PROTEIN.
GN P0436E04.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

```

OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.; genomic DNA, chromosome 1, PAC
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone:P0436E04.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002818; BAB16319.1; -.
SQ SEQUENCE 202 AA; 19763 MW; BFC2520037F8E274 CRC64;

Query Match      28.9%; Score 55.5; DB 10; Length 202;
Best Local Similarity 34.1%; Pred. No. 40;
Matches 14; Conservative 6; Mismatches 14; Indels 7; Gaps 1;

QY 1 IEQPTLRQWLAAAGGNGSGGIEGPTLRQWLAA 34
    :||| : ||| ||| :||| : ||| |||
Db 94 VVSPSCRQTAGRGCGGGRWMAAGRGDGGCRWWAA 134

RESULT 28
Q9UVX9 PRELIMINARY; PRT; 690 AA.
AC Q9UVX9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE GLUTAMINASE A.
GN GTAA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIB40;
RX MEDLINE=20406523; PubMed=10952006;
RA Koibuchi K., Nagasaki H., Yuasa A., Kataoka J., Kitamoto K.;
RT "Molecular cloning and characterization of a gene encoding glutaminase
RT from Aspergillus oryzae.";
RL Appl. Microbiol. Biotechnol. 54:59-68(2000).
DR EMBL: AB029552; BAA86934.1; -.
SQ SEQUENCE 690 AA; 76165 MW; E3D0B17841EEA0D CRC64;

Query Match      28.9%; Score 55.5; DB 3; Length 690;
Best Local Similarity 41.9%; Pred. No. 1.4e+02;
Matches 13; Conservative 4; Mismatches 9; Indels 5; Gaps 2;

QY 1 IEQPTLRQWLAAAGGNGSGGIEGPTLRQW 31
    :||| : ||| ||| :||| : |||
Db 36 VKSPYLSTWLSAGTDGNG-GYLAG---QW 61

RESULT 29
Q9HGS1 PRELIMINARY; PRT; 690 AA.
AC Q9HGS1;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE GLUTAMINASE (EC 3.5.1.2).
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U012;
RA Thammarongtham C., Turner G., Moir A.J., Tanticharoen M.,
RA Cheevadhanarak S.;
RT "A new class of glutaminase from Aspergillus oryzae.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

```


DR InterPro; IPR003101; KIX.
 DR InterPro; IPR000197; TAZ.finger.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02172; KIX; 1.
 DR Pfam; PF02135; zf-TAZ; 2.
 DR Pfam; PF00569; ZF; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; ZNF_ZZ; 1.
 DR SMART; SM00291; ZNF_ZZ; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
 DR PROSITE; PS0014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 3190 AA; 331879 MW; E53526F78BC055A8 CRC64;

Query Match 29.2%; Score 56; DB 5; Length 3190;
 Best Local Similarity 61.1%; Pred. No. 5.9e+02;
 Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGGIEGPT 27
 DB 44 LTGGAGGGGGGGASGVT 61

RESULT 24
 Q9W321 PRELIMINARY; PRT; 3275 AA.
 AC Q9W321; PRT; 3275 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG15319 PROTEIN.
 GN NEJ OR CG15319.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J., H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman F.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003448; AAF46516.1; -;
 DR FlyBase; FBgn0015624; nej.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003101; KIX.
 DR InterPro; IPR000197; TAZ.finger.
 DR InterPro; IPR000433; ZNF_ZZ.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02172; KIX; 1.
 DR Pfam; PF02135; zf-TAZ; 2.
 DR Pfam; PF00569; ZF; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; ZNF_ZZ; 1.
 DR SMART; SM00291; ZNF_ZZ; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS0014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 3275 AA; 340672 MW; E9944C3BFEC0E7AA CRC64;

Query Match 29.2%; Score 56; DB 5; Length 3275;
 Best Local Similarity 61.1%; Pred. No. 6e+02;
 Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGGIEGPT 27
 DB 44 LTGGAGGGGGGGASGVT 61

RESULT 25
 Q9QYX6 PRELIMINARY; PRT; 4833 AA.
 AC Q9QYX6; PRT; 4833 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ACZONIN.
 GN ACZ OR ACZ.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=9439764; PubMed=10508862;
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 RA Kilmann M.W.;
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rim and bassoon and binds
 RT profilin.";
 RL J. Cell Biol. 147:151-162(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX Kilmann M.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y19186; CAB60732.2; -;
 DR HSSP; P04410; 1A25.
 DR MGD; MGI:1349390; ACZ.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
 DR PROSITE; PS00004; C2_DOMAIN_2; 1.
 DR PROSITE; PS0106; PDZ; 1.
 SQ SEQUENCE 4833 AA; 525056 MW; FF22EE3AA1AF9F4F CRC64;

Wed Oct 9 10:30:17 2002

```

DE PUTATIVE CYTOCHROME P450 MONOOXYGENASE.
GN OSJNBAA001014.16.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartioideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Uteckback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.,
RA "Oryza sativa chromosome 10 BAC OSJNBAA001014 genomic sequence.";
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AC025783; AAK20054.1;
DR InterPro; IPR001128; Cyt_P450.
DR Pfam; PF00067; P450.1.
DR PRINTS; P00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 584 AA; 64797 MW; 1A55160A532DC83 CRC64;

Query Match 29.2%; Score 56; DB 10; Length 584;
Best Local Similarity 39.3%; Pred. No. 1e+02;
Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 6 LRWLAAAGGGGGGGGGTPTLQWLA 33
DB 36 LRPPSSGGGGGGAGGDEPPITTSWVS 63

RESULT 21
Q9HEA4 PRELIMINARY; PRT; 776 AA.
ID Q9HEA4
AC Q9HEA4
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN B11A5.200.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL451109; CAC18624.2;
KW Hypothetical protein.
SQ SEQUENCE 776 AA; 82771 MW; C9BEA870D94A37DE CRC64;

Query Match 29.2%; Score 56; DB 3; Length 776;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 13; Conservative 3; Mismatches 6; Indels 2;

QY 15 GGGNGSGGI---EG-PTLRQWLARA 36
DB 678 GGGGGGGVDDGDEPDPAGWLAAQA 703

RESULT 22
O14654 PRELIMINARY; PRT; 1257 AA.
ID O14654
AC O14654;

DE 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE INSULIN RECEPTOR SUBSTRATE 4 (DA24A23.2) (INSULIN RECEPTOR SUBSTRATE
DE 4).
GN DA24A23.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=97407931; PubMed=9261155;
RA Lavan B.E., Fantin V.R., Chang E.T., Lane W.S., Keller S.R.,
RA Lienhard G.E.;
RT "A novel 160-kDa phosphotyrosine protein in insulin-treated embryonic
RT kidney cells is a new member of the insulin receptor substrate
RT family.";
RL J. Biol. Chem. 272:21403-21407(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cobley V.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF007567; AAC51738.1;
DR EMBL; AL035425; CAB90290.1;
DR HSP; P35568; IIRS.
DR InterPro; IPR002404; Insulin_Recep_S-1.
DR InterPro; IPR001849; PH.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00169; PH; 1.
DR SMART; SM00628; INSULINRSI.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Receptor.
SQ SEQUENCE 1257 AA; 133766 MW; 4D512D65A7A80374 CRC64;

Query Match 29.2%; Score 56; DB 4; Length 1257;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 13 RAGGGNGSGGIEGP 26
DB 563 RFGGGHGGGGGGGP 576

RESULT 23
O01368 PRELIMINARY; PRT; 3190 AA.
ID O01368
AC O01368;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CREB-BINDING PROTEIN HOMOLOG.
GN NEJ OR CG15319.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97263578; PubMed=9109493;
RA Akimaru H., Chen Y., Dai P., Hou D.X., Nonaka M., Smolik S.M.,
RA Armstrong S., Goodman R.H., Ishii S.;
RT "Drosophila CBP is a co-activator of cubitus interruptus in hedgehog
RT signalling.";
RL Nature 386:735-738(1997).
DR EMBL; U88570; AAB53050.1;
DR TRANSFAC; T03236;
DR FlyBase; FBgn0015624; nejl.
DR InterPro; IPR001487; Bromodomain.

```

Dd 83 EGAAR-WRAAKSPARGQRGHRRRGGGCGGRPRRRR 123

RESULT 18
O80221 PRELIMINARY; PRT; 217 AA.

ID O80221;
AC O80221;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 24.3 KDA PROTEIN.
OS Methanobacterium phage psiM1.
DS Viruses.
OC NCBI_TaxID=78218;
OX RN [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99009353; PubMed=9791169;
RA Pfister P., Wasserfallen A., Stettler R., Leisinger M.
RT "Molecular analysis of Methanobacterium phage psiM2 Mol. Microbiol. 30:233-244(1998).
RL EMBL; AF065412; AAC27071.1; -
RS Hypothetical protein.
KW SEQUENCE 217 AA; 24325 MW; 9FG62B1BB5131468 CRC6C
SQ SEQUENCE 217 AA; 24325 MW; 9FG62B1BB5131468 CRC6C

Query Match 29.2%; Score 56; DB 9; Leng
Best Local Similarity 44.0%; Pred. No. 37;
Matches 11; Conservative 3; Mismatches 11; In

QY 1 IEGPTLRWLAAARGGNGSGTG 25
| : | : | : |||| | | |
Db 24 IDREALRDFLGISGGGGDGGSSG 48

RESULT 19
Q9FPB8 PRELIMINARY; PRT; 348 AA.

ID Q9FPB8;
AC Q9FPB8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE P0679C08.26 PROTEIN.
DN P0679C08.26.
OS Eryza sativa (rice)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Po
OC Ehrhartoidae; Orzyaceae; Orzya.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cv. NIPPONARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oriza sativa nipponbare(GA3) genomic DNA, chromosom
clone:P0679C08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databa
EMBL; AP002542; BAB19386.1; -
DS SEQUENCE 348 AA; 37176 MW; DAABE5D080883E94C CRC6
SQ SEQUENCE 348 AA; 37176 MW; DAABE5D080883E94C CRC6

Query Match 29.2%; Score 56; DB 10; Leng
Best Local Similarity 37.9%; Pred. No. 61;
Matches 11; Conservative 7; Mismatches 9; Inc

QY 5 TLROWLAARA--GGGNGSGTGPTLRW 31
| : | : | : | | | | | | |
Db 85 TIRAFLESLPPGSGGDGPTPSVAW 113

RESULT 20
Q9AV27 PRELIMINARY; PRT; 584 AA.

ID Q9AV27;
AC Q9AV27;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Richards S., Ashburner M., Henderson S.N.,
 RA George R.A., Lewis S.E., Siedel M.D., Zhang Q., Chen L.X.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Pfeiffer B.D.,
 RA Rondon R.C., Rogers Y.-H.C., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley R.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003464; AAF47228.1;
 DR FlyBase; FBgn0035024; CG11414.
 DR InterPro; IPR000822; Znf-C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR DNA-binding; Metal-binding; Zinc-finger.
 KW SEQUENCE 867 AA; 95938 MW; F4998152BB02D18C CRC64;
 QY 15 GGGNGSGGIEGPTLRQWLAA 36
 DB 401 GGGNG-GGVTGPTSSS-IAA 420
 Query Match 29.7%; Score 57; DB 5; Length 867;
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;
 Matches 14; Conservative 2; Mismatches 4; Indels 2; Gaps 2;
 RESULT 14
 Q9A535 PRELIMINARY; PRT; 249 AA.
 AC Q9A535;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE ABC TRANSPORTER, ATP-BINDING PROTEIN.
 GN CC2634.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CBJ5;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Urtuback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005931; AAK24601.1;
 DR TIGR; CC2634;
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR001687; ATP_GTP_A.
 DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR ATP-binding; Complete proteome.
 KW SEQUENCE 249 AA; 25448 MW; AD63BD04C1F77122 CRC64;
 Query Match 29.4%; Score 56.5; DB 16; Length 249;
 Best Local Similarity 42.9%; Pred. No. 38;
 Matches 15; Conservative 5; Mismatches 14; Indels 1; Gaps 1;
 QY 1 IEGPTLRQWLAAAGGSGGIEGPTLRQWLAA 35
 DB 54 LEQPT-RGAVSRLLAGKGETSVVFQPTLRQWLAA 87
 RESULT 15
 Q9CCCC PRELIMINARY; PRT; 488 AA.
 ID Q9CCCC;
 AC Q9CCCC;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE POSSIBLE ATP/GTP-BINDING PROTEIN.
 GN ML0997.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL583920; CAC31376.1;
 DR Leproma; ML0997;
 DR InterPro; IPR000765; GTP1_OBG.
 DR PRINTS; PR00326; GTP1OBG.
 DR Complete proteome.

1. The first part of the document is a list of the names of the persons who have been appointed to the various positions of the Board of Directors of the company. The names are listed in alphabetical order, and each name is followed by the position to which the person has been appointed. The names are: Mr. John A. Smith, President; Mr. James B. Jones, Vice President; Mr. Robert C. Brown, Secretary; Mr. William D. White, Treasurer; Mr. Charles E. Black, Chairman of the Board; Mr. Thomas F. Green, Director; Mr. Richard H. Hall, Director; Mr. Daniel I. King, Director; Mr. Benjamin L. Lee, Director; Mr. Samuel M. Miller, Director; Mr. David N. Moore, Director; Mr. Joseph O. Parker, Director; Mr. Henry P. Reed, Director; Mr. George Q. Ross, Director; Mr. Frank R. Taylor, Director; Mr. Edward S. Turner, Director; Mr. John T. Walker, Director; Mr. William U. Young, Director.

KW Hypothetical protein.
SQ SEQUENCE 439 AA; 47297 MW; 533EPC240CEA1BA2 CRC64;

Query Match 30.2%; Score 58; DB 10; Length 439;
Best Local Similarity 30.0%; Pred. No. 45; Indels 14; Gaps 1;
Matches 15; Conservative 2; Mismatches 19

QY 1 TEGPTLRQLARAGGNGSGG-----TEGPTLRQLARLA 36
| | | | | | | | | | | | | | | | | | | |
Db 39 LHAPLLRLWPLGGGGGGGGGGGGGAVRGVGEARSORAAEA 88
| | | | | | | | | | | | | | | | | | | |

RESULT 8
Q9ZRB9 PRELIMINARY; PRT; 392 AA.

ID Q9ZRB9
AC Q9ZRB9
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL TRANSMEMBRANE PROTEIN (FRAGMENT).
GN THOX1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP+, 93-137;
RA Parnis A., Lifschitz E.;
RT "The Tomato Homeobox 1 gene."
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U76409; AAC09582.1; -.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; UNKNOWN_1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
FT NON_TER 1
SQ SEQUENCE 392 AA; 43306 MW; C18916A988063DA8 CRC64;

Query Match 29.9%; Score 57.5; DB 10; Length 392;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 12; Conservative 0; Mismatches 3; Indels 3; Gaps 1

QY 8 QWL---AARAGGNGSGG 22
| | | | | | | | | | | | | | | | | | | |
Db 58 QWLSPTAAAGGNGSGG 75
| | | | | | | | | | | | | | | | | | | |

RESULT 9
Q92KH3 PRELIMINARY; PRT; 928 AA.

ID Q92KH3
AC Q92KH3
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL TRANSMEMBRANE PROTEIN SMC02633.
GN SMC02633.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Bailey-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA

RESULT 4					
ID	O83436	PRELIMINARY;	PRT;	683 AA.	
AC	Q943K0				
DT	01-NOV-1998	(TREMBLrel. 08, Created)			
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	CONSERVED HYPOTHETICAL PROTEIN.				
DE	TP0421.				
GN	Treponema pallidum.				
OC	Bacteris; Spirochaetales; Spirochaetaceae; Treponema.				
OX	NCBI_TaxID=160;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NICHOLS;				
RC	MEDLINE=98332770; PubMed=9665876;				
RA	Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,				
RA	Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,				
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,				
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,				
RA	McDonald L., Artisch P., Bowman C., Cotton M.D., Fujii C., Garland S.,				
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,				
RA	Venter J.C.;				
RT	*Complete genome sequence of Treponema pallidum, the syphilis				
RT	spirochete.*;				
RL	Science 281:375-388(1998).				
DR	EMBL; AE001220; AAC65409.1; -				
DR	TIGR; TP0421; -				
DR	InterPro; IPR001258; NHL.				
DR	InterPro; IPR001440; TPR.				
DR	Pfam; PF01436; NHL; 4.				
DR	Pfam; PF00515; TPR; 1.				
KW	Complete proteome.				
KW	SEQUENCE 683 AA; 74518 MW; F91407FA7094AAD1 CRC64;				
Query Match	31.2%; Score 60; DB 16; Length 683;				
Best Local Similarity	43.8%; Pred.No.42;				
Matches	14; Conservative	2; Mismatches	12; Indels	4; Gaps	
OY	4	PTLRQLAARAGGNGSGIEGPTLRQLAAR 35			
DB	74	PLILEWL----GNAYRSGIEGAALHQWGAA 101			
	: :				
RESULT 5					
O943K0	PRELIMINARY;	PRT;	253 AA.		
ID	Q943K0				
AC	Q943K0;				
DT	01-DEC-2001	(TREMBLrel. 19, Created)			
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	P0039A07.6	PROTEIN.			
GN	P0039A07.6.				
OS	Oryza sativa (Rice).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
OC	Ericaceae; Oryzaceae; Oryza.				
OX	NCBI_TaxID=4530;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. NIPPONBARE;				
RA	Sasaki T., Matsumoto T., Yamamoto K.;				
RT	"Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC				
RT	clone:P0039A07.";				
RL	Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF003235; BAB64100.1; -				
DR	SEQUENCE 253 AA; 25568 MW; A96316CE5F97B2B CRC64;				
Query Match	30.7%; Score 59; DB 10; Length 253;				
Best Local Similarity	48.1%; Pred.No.20;				
Matches	13; Conservative	3; Mismatches	11; Indels	0; Gaps	

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-32
Perfect score: 192
Sequence: 1 IEPTLROWLAARAGGNGGGIEPTLROWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	62	32.3	360	10 Q9LGC9	Q9LGC9 oryza sativ
2	61.5	32.0	431	13 Q9PVG9	Q9PVG9 coturnix co
3	60	31.2	491	10 Q94LP1	Q94LP1 oryza sativ
4	60	31.2	683	16 O83436	O83436 treponema p
5	59	30.7	253	10 Q943K0	Q943K0 oryza sativ
6	59	30.7	496	2 Q9AD76	Q9AD76 streptomyce
7	58	30.2	439	10 Q9SDK6	Q9SDK6 oryza sativ
8	57.5	29.9	392	10 Q92RB9	Q92RB9 lycopersico
9	57.5	29.9	928	16 Q92KH3	Q92KH3 rhizobium m
10	57	29.7	200	10 Q9ST53	Q9ST53 arabidopsis
11	57	29.7	500	5 Q19476	Q19476 caenorhabdi
12	57	29.7	607	2 Q9L8D4	Q9L8D4 polyangium
13	57	29.7	867	5 Q9W149	Q9W149 drosophila
14	56.5	29.4	249	16 Q9A535	Q9A535 caulobacter
15	56.5	29.4	488	16 Q9CCC0	Q9CCC0 mycobacteri
16	56.5	29.4	518	2 Q49843	Q49843 mycobacteri

17	56	29.2	125	10 Q9LWC8	Q9LWC8 oryza sativ
18	56	29.2	217	9 O80221	O80221 methanobact
19	56	29.2	348	10 Q9FPB8	Q9FPB8 oryza sativ
20	56	29.2	584	10 Q9AV27	Q9AV27 oryza sativ
21	56	29.2	776	3 Q9HEA4	Q9HEA4 neurospora
22	56	29.2	1257	4 O14654	O14654 homo sapien
23	56	29.2	3190	5 O01368	O01368 drosophila
24	56	29.2	3275	5 Q9W321	Q9W321 drosophila
25	56	29.2	4833	11 Q9QYX6	Q9QYX6 mus musculu
26	56	29.2	5038	11 Q9QYX7	Q9QYX7 mus musculu
27	55.5	28.9	202	10 Q9FTZ5	Q9FTZ5 oryza sativ
28	55.5	28.9	690	3 Q9UVX9	Q9UVX9 aspergillus
29	55.5	28.9	690	3 Q9HGS1	Q9HGS1 aspergillus
30	55	28.6	103	5 Q95QV6	Q95QV6 caenorhabdi
31	55	28.6	134	10 Q9ARU5	Q9ARU5 oryza sativ
32	55	28.6	164	10 Q9M6A0	Q9M6A0 catharanthu
33	55	28.6	170	5 Q9W033	Q9W033 drosophila
34	55	28.6	545	16 Q9ABY5	Q9ABY5 caulobacter
35	55	28.6	900	5 Q9VK09	Q9VK09 drosophila
36	54.5	28.4	209	16 Q9A470	Q9A470 caulobacter
37	54.5	28.4	495	16 O33230	O33230 mycobacteri
38	54.5	28.4	869	5 Q9VZ82	Q9VZ82 drosophila
39	54.5	28.4	990	10 Q9FWK7	Q9FWK7 oryza sativ
40	54.5	28.4	1431	11 Q9JMH4	Q9JMH4 mesocricetu
41	54	28.1	377	13 Q9VHD0	Q9VHD0 petromyzon
42	54	28.1	509	2 Q9S5E5	Q9S5E5 streptomyce
43	54	28.1	529	10 Q9ASE5	Q9ASE5 oryza sativ
44	54	28.1	612	4 Q9P270	Q9P270 homo sapien
45	54	28.1	640	10 Q96397	Q96397 chlamydomon

ALIGNMENTS

RESULT 1

Q9LGC9 ID Q9LGC9 PRELIMINARY; PRT; 360 AA.
 AC Q9LGC9;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DF 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE PUTATIVE ZINC FINGER PROTEIN.
 GN P0462H08.19.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzaeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone: P0462H08.19";
 RL EMBL: AP002525; BAB07996.1;
 DR Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR InterPro: IPR000571; Zf-CCCH.
 DR Pfam: PF00842; Zf-CCCH; 4.
 DR SMART: SM00356; Znf_C3H1; 4.
 SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 32.3%; Score 62; DB 10; Length 360;
 Best Local Similarity 48.0%; Pred. No. 13;
 Matches 12; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 IEPTLROWLAARAGGNGGGIEG 25

Db 26 LEGPMWRMGLGGGGGGGGGGDG 50

RESULT 2

Q9PVG9 ID Q9PVG9 PRELIMINARY; PRT; 431 AA.


```
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS00002; SH3; 1.
DR SH3 domain; Membrane; Calmodulin-binding; Transferase;
KW Phosphorylation; ATP-binding; Serine/Threonine-protein kinase.
FT DOMAIN 12 278 CALMODULIN-BINDING.
FT DOMAIN 306 316 PDZ.
FT DOMAIN 493 568 SH3.
FT DOMAIN 583 650 SH3.
FT DOMAIN 710 897 GUANYLATE KINASE.
FT NP-BIND 18 26 ATP (BY SIMILARITY).
FT BINDING 41 41 ATP (BY SIMILARITY).
FT ACT_SITE 141 141 BY SIMILARITY.
FT MOD_RES 292 292 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 7 7 V -> L (IN REF. 2).
FT CONFLICT 136 136 E -> D (IN REF. 2).
FT CONFLICT 463 463 MISSING (IN REF. 2).
FT CONFLICT 496 496 H -> Q (IN REF. 2).
FT CONFLICT 522 522 L -> I (IN REF. 2).
FT CONFLICT 631 631 G -> D (IN REF. 2).
FT CONFLICT 667 667 MISSING (IN REF. 2).
FT CONFLICT 719 719 H -> HH (IN REF. 2).
FT CONFLICT 739 739 V -> A (IN REF. 2).
SQ SEQUENCE 897 AA; 100792 MW; FFE88E3F5D2035BE CRC64;

Query Match 26.3%; Score 50.5; DB 1; Length 897;
Best Local Similarity 44.4%; Pred. No. 1.8e+02;
Matches 12; Conservative 2; Mismatches 12; Indels 1; Gaps 1;

Qy 9 WLAARAGGNGSGG-TEGPTLRWLAA 34
| | | | | | | | | | | | | | |
Db 625 WWOARLGTGSGAGLIPSPQLQEWRIA 651

RESULT 29
LYCV_BPP21
ID LYCV_BPP21 STANDARD; PRT; 165 AA.
AC P27359;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysozyme (EC 3.2.1.17) (Lysis protein) (Muramidase) (Endolysin).
GN R.
OS Bacteriophage P21 (Bacteriophage 21).
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10711;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91210180; PubMed=2019562;
RA Bonovich M.T., Young R.;
RT "Dual start motif in two lambdaoid S genes unrelated to lambda S.";
RL J. Bacteriol. 173:2897-2905(1991).
CC -!- FUNCTION: ESSENTIAL FOR LYSIS OF BACTERIAL CELL WALL, BY SHOWING
CC CELL WALL HYDROLYZING ACTIVITY. ACTS AS A TRANSGLYCOSYLASE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -!- SIMILARITY: BELONGS TO FAMILY 24 OF GLYCOSYL HYDROLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J02580; AAA32300.1; -.
DR PIR; C25647; WMBPP2.
DR InterPro; IPR002196; Phage_lysozyme.
DR Pfam; PF00959; Phage_lysozyme; 1.
KW Late protein; Hydrolase; Glycosidase; Bacteriolytic enzyme.
FT ACT_SITE 35 35 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 44 44 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 44 44 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 165 AA; 17998 MW; 2DA65CCF1EA9329B CRC64;

Query Match 26.0%; Score 50; DB 1; Length 165;
Best Local Similarity 35.3%; Pred. No. 43;
Matches 12; Conservative 6; Mismatches 4; Indels 12; Gaps 1;

Qy 4 PTLROWLAARAGG-----NGSGGIEG 25
| | | | | | | | | | | | | | |
Db 3 PSLRKAVAAAIAGGGAIAIASVLITGPSGNDGLEG 36

RESULT 30
LYCV_BPPA2
ID LYCV_BPPA2 STANDARD; PRT; 165 AA.
AC P10439;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysozyme (EC 3.2.1.17) (Lysis protein) (Muramidase) (Endolysin).
GN 15.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=10738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304457; PubMed=3017988;
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and mmpC outer membrane porin protein genes of
RT lambdaoid bacteriophage.";
RL J. Biol. Chem. 261:12723-12732(1986).
CC -!- FUNCTION: ESSENTIAL FOR LYSIS OF BACTERIAL CELL WALL, BY SHOWING
CC CELL WALL HYDROLYZING ACTIVITY. ACTS AS A TRANSGLYCOSYLASE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -!- SIMILARITY: BELONGS TO FAMILY 24 OF GLYCOSYL HYDROLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J02580; AAA32300.1; -.
DR PIR; C25647; WMBPP2.
DR InterPro; IPR002196; Phage_lysozyme.
DR Pfam; PF00959; Phage_lysozyme; 1.
KW Late protein; Hydrolase; Glycosidase; Bacteriolytic enzyme.
FT ACT_SITE 35 35 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 44 44 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 44 44 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 165 AA; 17998 MW; 2DA65CCF1EA9329B CRC64;

Query Match 26.0%; Score 50; DB 1; Length 165;
Best Local Similarity 35.3%; Pred. No. 43;
Matches 12; Conservative 6; Mismatches 4; Indels 12; Gaps 1;

Qy 4 PTLROWLAARAGG-----NGSGGIEG 25
| | | | | | | | | | | | | | |
Db 3 PSLRKAVAAAIAGGGAIAIASVLITGPSGNDGLEG 36

Search completed: October 9, 2002, 09:00:20
Job time : 5.3831 secs
```

DR Pfam: PF00037; fer4; 3.
 DR Pfam: PF01077; NIR_SIR; 1.
 DR PRINTS: PR00397; SROHAEM.
 DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
 DR PROSITE: PS00365; NIR_SIR; 1.
 KW Hypothetical protein; Oxidoreductase; Heme; Iron-sulfur; 4Fe-4S;
 KW Complete proteome.
 FT METAL 428 428 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 434 434 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 468 468 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 472 472 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT (BY SIMILARITY).
 SQ SEQUENCE 620 AA; 69793 MW; 9D71D2580D7D0BA8 CRC64;
 Query Match 26.3%; Score 50.5; DB 1; Length 620;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;
 QY 2 EGPTLROWLAARAGGNGSGGIEGPT 27
 DB 418 EGPLVRATLAC-PGCGNCSSGLVDIT 442
 RESULT 28
 CAKI_DROME STANDARD; PRT; 897 AA.
 AC Q2410; Q24272; Q9VD77; Q9VD79;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calcium/calmodulin-dependent protein kinase (EC 2.7.1.123) (CAKI)
 DE (Peripheral plasma membrane protein CAMGUK).
 GN CAKI OR CMG OR CGL3412 OR CG6703
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=97321552; PubMed=9178262;
 RA Dimitrakos S.D., Woods D.F., Bryant P.J.;
 RT "Camguk, lin-2, and CASK: novel membrane-associated guanylate kinase
 RT homologs that also contain Cam kinase domains.";
 RL Mech. Dev. 63:127-130(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=96203108; PubMed=8617233;
 RA Martin J.-R., Ollo R.;
 RT "A new Drosophila Ca2+/calmodulin-dependent protein kinase (Caki) is
 RT localized in the central nervous system and implicated in walking
 RT speed.";
 RL EMBO J. 15:1865-1876(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananthakrishnan P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brookstein P., Brotter P.,
 RA Burtis K.C., Busan M.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Posler C., Gabrielian A.E., Garb N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobary C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Wang X.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster.";
 Science 287:2185-2195(2000).
 RL - FUNCTION: MAY REGULATE TRANSMEMBRANE PROTEINS THAT BIND CALCIUM,
 CC CALMODULIN, OR NUCLEOTIDES.
 CC - CATALYTIC ACTIVITY: ATP + protein = ADP + O-phosphoprotein.
 CC - TISSUE SPECIFICITY: DURING EMBRYOGENESIS, LARVAL AND PUPAL LIFE,
 CC FOUND ALMOST EXCLUSIVELY IN THE CENTRAL NERVOUS SYSTEM. IN THE
 CC ADULT HEAD FOUND IN THE LAMINA, THE NEUROPIIL OF THE MEDULLA,
 CC LOBULA, LOBULA PLATE AND IN THE CENTRAL BRAIN.
 CC - SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE SER/THR
 CC FAMILY OF PROTEIN KINASES. CAMK SUBFAMILY.
 CC - SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC - SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC - CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
 CC FRAMESHIFTS AND REF.3 SEQUENCE DIFFERS DUE TO ONE FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U53190; AAC80169.1; --
 CC EMBL; X94264; CAA63940.1; ALT_FRAME.
 CC EMBL; AE003736; AAF55920.1; ALT_SEQ.
 CC EMBL; AE003736; AAF55922.1; ALT_SEQ.
 CC HSP; O14936; IKWA.
 CC FlyBase; FBgn0013759; Caki.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR004172; L27.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.
 CC InterPro; IPR002290; Ser_thr_pkinase.
 CC Pfam; PF00625; Guanylate_kin; 1.
 CC Pfam; PF02828; L27; 2.
 CC Pfam; PF00599; pkinase; 1.
 CC Pfam; PF00018; SH3; 1.
 CC Pfam; PF00017; GuKc; 1.
 CC SMART; SM00072; Gukc; 1.
 CC SMART; SM00228; PDZ; 1.
 CC SMART; SM00326; SH3; 1.
 CC SMART; SM00220; S_TKc; 1.
 CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 CC PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
 CC PROSITE; PS50106; PDZ; 1.

```

DR PIR; S22765; S22765.
DR MIM; 602869;
DR InterPro; IPR003034; SAP.
DR InterPro; IPR003877; SPRY.
DR InterPro; IPR003878; SPRY_domain.
DR Pfam; PF02037; SAP; 1.
DR Pfam; PF00822; SPRY; 1.
DR SMART; SM00513; SAP; 1.
DR SMART; SM00449; SPRY; 1.
DR KW Nuclear protein; Ribonucleoprotein; RNA-binding; DNA-binding;
KW Phosphorylation; ATP-binding; Alternative splicing.
FT DOMAIN 1 160 ASP/OLIGO-RICH (ACIDIC).
FT DOMAIN 161 209 GIN-RICH.
FT DOMAIN 702 824 GLY-RICH.
FT DOMAIN 84 94 POLY-GLU.
FT NP_BIND 503 510 ATP (POTENTIAL).
FT DOMAIN 713 738 RNA-BINDING (RGG-BOX).
FT DOMAIN 739 749 POLY-GLY.
FT VARSPLIC 213 230 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 824 AA; 90479 MW; F7D04BEA48C8FEC CRC64;

Query Match 26.68; Score 51; DB 1; Length 824;
Best Local Similarity 71.44; Pred. No. 1.5e+02;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 13 RAGGGGGGGGIEGP 26
I I I I I I I I I I
Db 738 RGGGGGGGGGIGYP 751

RESULT 26
SPML_RAT
ID SPML_RAT STANDARD; PRT; 335 AA.
AC P56225;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sperm 1 POU-domain transcription factor (SPRM-1).
GN SPML
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=94068549; PubMed=7902581;
RA Andersen B., Fearse R.V. II, Schlegel P.N., Cichon Z.,
RA Schonemann M.D., Bardin C.W., Rosenfeld M.G.;
RT "Sperm 1: a POU-domain gene transiently expressed immediately before
RT meiosis I in the male germ cell."
RL Proc. Natl. Acad. Sci. U.S.A. 90:11084-11088(1993).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
CC OCTAMER MOTIF ('ATGTTAAT'). MAY EXERT A REGULATORY FUNCTION IN
CC MEIOTIC EVENTS THAT ARE REQUIRED FOR TERMINAL DIFFERENTIATION OF
CC MALE GERM CELL.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: HIGHLY RESTRICTED TO ADULT TESTIS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED TRANSIENTLY IMMEDIATELY PRIOR TO
CC MEIOSIS I DURING SPERMATOGENESIS IN THE MALE GERM CELL.
CC -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L23864; -; NOT_ANNOTATED_CDS.
CC HSSP; P20263; IOCp.
CC InterPro; IPR001356; Homeobox.

```

```

DR InterPro; IPR000327; POU.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF00157; pou; 1.
DR PRINTS; PO00028; POU_DOMAIN.
DR ProDom; PD000583; POU; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00352; POU; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein;
KW Spermatogenesis; Meiosis.
FT DOMAIN 117 187 POU.
FT DNA_BIND 205 264 HOMEBOX.
SQ SEQUENCE 335 AA; 37159 MW; 820B9AC12D6B10D7 CRC64;

Query Match 26.38; Score 50.5; DB 1; Length 335;
Best Local Similarity 40.08; Pred. No. 72;
Matches 10; Conservative 5; Mismatches 5; Indels 5; Gaps 1;

QY 17 GNSGGIEGPTLRQ-----WLAARA 36
I I I I I I I I I I I I I I I I
Db 14 GNSGGLEGPPVPMRVDPMTWLSQA 38

RESULT 27
Y870_METJA
ID Y870_METJA STANDARD; PRT; 620 AA.
AC Q58280;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0870.
GN MJ0870.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO COENZYME F420 HYDROGENASE BETA SUBUNIT.
CC -!- SIMILARITY: TO M. JANNASCHII MJ1349, MJ0725 AND MJ0551.
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67531; AAB98876.1; -
CC HSSP; Q45560; 1BQX.
CC TIGR; MJ0870; -
CC InterPro; IPR001450; 4Fe4S_ferredoxin.
CC InterPro; IPR000660; Nif_Sif.

```

Db 453 GPPRRGGHAQKLGSGRTGQMEG 475

FT REPEAT 479 512 TPR 8.
FT REPEAT 514 547 TPR 10
FT REPEAT 548 581 TPR 10
SQ SEQUENCE 611 AA; 67521 MW; 486fB79FC4CE5BAF CRC64;

Query Match 26.6%; Score 51; DB 1; Length 611;
Best Local Similarity 44.1%; Pred. No. 1.le-02;
Matches 15; Conservative 1; Mismatches 10; Indels 2;

RESULT 24
OM70_MOUSE STANDARD; PRT; 611 AA.

ID OM70_MOUSE
AC Q9CZM5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitochondrial precursor proteins import receptor (Translocase of outer membrane TOM70).
GN TOM70A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Embryo;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa K., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasakawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo T., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection."
Nature 409:685-690(2001).
CC -!- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL
CC -!- MITOCHONDRIAL PRECURSOR PROTEINS (BY SIMILARITY).
CC -!- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC LEAST 8 DIFFERENT PROTEINS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC outer membrane (By similarity).
CC -!- SIMILARITY: CONTAINS 10 TPR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL: AK012084; BAB28018.1; --
CC MGD: MG1_106295; D16wsu109e.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 10.
DR SMART: SM00028; TPR; 10.
KW Mitochondrion; Outer membrane; Transmembrane; Repeat; TPR repeat.
FT DOMAIN 1 9 INTERMEMBRANE (POTENTIAL).
FT TRANSMEM 10 30 POTENTIAL.
FT DOMAIN 31 611 CYTOPLASMIC (POTENTIAL).
FT REPEAT 117 150 TPR 1.
FT REPEAT 156 189 TPR 2.
FT REPEAT 297 330 TPR 3.
FT REPEAT 332 365 TPR 4.
FT REPEAT 370 403 TPR 5.
FT REPEAT 404 437 TPR 6.
FT REPEAT 445 478 TPR 7.

Query Match 26.8%; Score 51.5; DB 1; Length 619;
Best Local Similarity 50.08; Pred. No. 99;
Matches 13; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 7 ROWLAARAGGGGGGEGPTLRQ-W 31
DB 39 RQDSQAERYGGGGGCGNPTNRQW 64

RESULT 22

YR21_TRSVR STANDARD; PRT; 201 AA.
AC P25245;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 16-Oct-2001 (Rel. 40, Last annotation update)
DE Hypothetical 20.2 kDa protein in RNA2.
OS Tomato ringspot virus (isolate raspberry) (Tomrsv).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus
OX NCBI_TaxID=12281;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91311402; PubMed=1856689;
RA Rott M.E., Tremaine J.H., Rochon D.M.;
RT "Nucleotide sequence of tomato ringspot virus RNA-2.";
RL J. Gen. Virol. 72:1505-1514(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D12477; BAA02044.1; -
DR PIR; JQ1094;
DR HSSP; P04002; LWFA.
KW Hypothetical protein.
FT DOMAIN 15 22 POLY-GLY.
FT DOMAIN 61 66 POLY-GLY.
FT DOMAIN 144 148 POLY-GLY.
SQ SEQUENCE 201 AA; 20194 MW; 903850618D7B450 CRC64;

Query Match 26.68; Score 51; DB 1; Length 201;
Best Local Similarity 53.8%; Pred. No. 40;
Matches 14; Conservative 1; Mismatches 7; Indels 4; Gaps 1;

QY 13 RAGGGNGGGGIE----GPTLRQWLA 34
DB 13 RAGGGGGGGGKEVFKAGRTLLKVLKA 38

RESULT 23

ID G3B2_HUMAN STANDARD; PRT; 482 AA.
AC Q9UN86; O75149; O60506;
DT 16-Oct-2001 (Rel. 40, Created)
DT 16-Oct-2001 (Rel. 40, Last sequence update)
DE Ras-Grpase-activating protein binding protein 2 (GAP SH3-domain
DE binding protein 2) (G3BP-2).
GN G3BP2 OR KIAA0660.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE=Brain;
RA Kennedy D., Mattick J.S.

*Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b,
members of a novel SH3 domain-binding and RNA-binding protein family
implicated in signal transduction.";
Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

[2]
RN SEQUENCE FROM N.A. (ISOFORM A).
RP TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).

[3]
RN SEQUENCE FROM N.A. (ISOFORM B).
RP TISSUE=Brain;
RX Guillard E.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

[4]
RN INTERACTION WITH IKAPPABALPHA.
RP MEDLINE=20549669; PubMed=10969074;
RX Prigent M., Barlat I., Langen H., Dargemont C.;
RT "Ikappabalpha and Ikappabalpha /NF-kappa B complexes are retained in
the cytoplasm through interaction with a novel partner, RasGAP
SH3-binding protein 2.";
RL J. Biol. Chem. 275:36441-36449(2000).

CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA
CC TRANSPORT (POTENTIAL).

CC -1- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF145284; AAD51932.1; -
DR EMBL; AB014560; BAA31635.1; -
DR EMBL; AF051311; AAC15705.1; -
DR HSSP; P09651; 1HAL.

DR InterPro; IPR002075; NTF2.

DR InterPro; IPR000504; RRM.

DR Pfam; PF02136; NTF2; 1.

DR Pfam; PF00076; rrm; 1.

DR SMART; SM00360; RRM; 1.

DR PROSITE; PS50177; NTF2_DOMAIN; 1.

DR PROSITE; PS50102; RRM; 1.

DR PROSITE; PS0030; RRM_RNP_1; FALSE_NEG.

KW Transport; RNA-binding; Alternative splicing.

FT DOMAIN 11 133 NTF2.

FT DOMAIN 331 409 RNA-BINDING (RRM).

FT DOMAIN 134 223 GLU-RICH.

FT DOMAIN 419 479 GLY-RICH.

FT VARSPLIC 243 275 MISSING (IN ISOFORM B).

FT CONFLICT 267 267 S -> P (IN REF. 2).

FT CONFLICT 359 359 E -> V (IN REF. 3).

FT CONFLICT 460 460 M -> I (IN REF. 3).

SQ SEQUENCE 482 AA; 54111 MW; 2545C6A3F1AAE218 CRC64;

Query Match 26.6%; Score 51; DB 1; Length 482;
Best Local Similarity 43.5%; Pred. No. 89;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGGGNGGGGIEG 25

|| | | | | | | | | | |

RA Germann U.A., Lerch K.;
 RT "Isolation and partial nucleotide sequence of the laccase gene from
 RT Neurospora crassa: amino acid sequence homology of the protein to
 RT human ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).
 CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
 CC H(2)O.
 CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M14554; AAA33590.1; -;
 CC EMBL; M18333; AAA33591.1; -;
 CC PIR; A28523; KSNCLD.
 CC PIR; A29762; A29762.
 CC InterPro; IPR001117; Cu-oxidase.
 CC InterPro; IPR002355; Multicopperoxidase2.
 CC Pfam; PF00394; Cu-oxidase; 3.
 CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
 CC PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 CC Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 KW Glycoprotein; Repeat.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 49 LACCASE.
 FT CHAIN 50 606 PLASTOCYANIN-LIKE 1.
 FT PROPEP 607 619 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 216 373 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 431 566
 FT METAL 144 146 COPPER (TYPE 2) (PROBABLE).
 FT METAL 146 146 COPPER (TYPE 3) (PROBABLE).
 FT METAL 189 189 COPPER (TYPE 3) (PROBABLE).
 FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
 FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
 FT METAL 477 477 COPPER (TYPE 1) (PROBABLE).
 FT METAL 480 480 COPPER (TYPE 2) (PROBABLE).
 FT METAL 482 482 COPPER (TYPE 3) (PROBABLE).
 FT METAL 548 548 COPPER (TYPE 3) (PROBABLE).
 FT METAL 549 549 COPPER (TYPE 1) (PROBABLE).
 FT METAL 550 550 COPPER (TYPE 3) (PROBABLE).
 FT METAL 554 554 COPPER (TYPE 3) (PROBABLE).
 FT METAL 559 559 COPPER (TYPE 1) (PROBABLE).
 FT METAL 559 559 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 619 AA; 68198 MW; FDED6D7BB65048E3 CRC64;
 Query Match 26.8%; Score 51.5; DB 1; Length 619;
 Best Local Similarity 50.0%; Pred. No. 99;
 Matches 13; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
 QY 7 ROWLAARAGGNGSGIGPTLRQ-W 31
 || | || | || | || |
 Db 39 RODSQAERYGGGGGCGCNSPTNRQGW 64
 RESULT 21
 LAC2_NEUCR

ID LAC2_NEUCR STANDARD; PRT; 619 AA.
 AC P10574;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
 DE (Urishiol oxidase) (Laccase allele TS).
 GN LACC.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP MEDLINE=88087214; PubMed=2961749;
 RX Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
 RT "Characterization of two allelic forms of Neurospora crassa laccase.
 RT Amino- and carboxyl-terminal processing of a precursor.";
 RL J. Biol. Chem. 263:885-896(1988).
 CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
 CC H(2)O.
 CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M18334; AAA33592.1; -;
 CC PIR; B28523; KSNCLT.
 CC InterPro; IPR001117; Cu-oxidase.
 CC InterPro; IPR002355; Multicopperoxidase2.
 CC Pfam; PF00394; Cu-oxidase; 3.
 CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
 CC PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 CC Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 KW Glycoprotein; Repeat.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 49 LACCASE.
 FT CHAIN 50 606 PLASTOCYANIN-LIKE 1.
 FT PROPEP 607 619 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 216 373 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 431 566
 FT METAL 144 144 COPPER (TYPE 2) (PROBABLE).
 FT METAL 146 146 COPPER (TYPE 3) (PROBABLE).
 FT METAL 189 189 COPPER (TYPE 3) (PROBABLE).
 FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
 FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
 FT METAL 477 477 COPPER (TYPE 1) (PROBABLE).
 FT METAL 480 480 COPPER (TYPE 2) (PROBABLE).
 FT METAL 482 482 COPPER (TYPE 3) (PROBABLE).
 FT METAL 548 548 COPPER (TYPE 3) (PROBABLE).
 FT METAL 549 549 COPPER (TYPE 1) (PROBABLE).
 FT METAL 550 550 COPPER (TYPE 3) (PROBABLE).
 FT METAL 554 554 COPPER (TYPE 3) (PROBABLE).
 FT METAL 559 559 COPPER (TYPE 1) (PROBABLE).
 FT METAL 559 559 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 619 AA; 68120 MW; 0BB6CCDE18841145 CRC64;

Wed Oct 9 10:30:16 2002

```

DR EMBL; M33512; AAA35586.1; JOINED.
DR EMBL; Z15025; CAA78744.1; -.
DR PIR; B35098; B35098.
DR PIR; S36152; S36152.
DR MIM; 142580; -.
KW Repeat.
FT DOMAIN 519 524 POLY-PRO.
FT DOMAIN 636 657 GLN-RICH.
FT DOMAIN 684 688 POLY-PRO.
FT DOMAIN 699 704 POLY-PRO.
FT DOMAIN 814 821 POLY-PRO.
FT DOMAIN 814 821 POLY-PRO.
FT DOMAIN 1340 1345 POLY-GLY.
FT DOMAIN 1398 1403 POLY-GLY.
FT DOMAIN 1436 1442 POLY-PRO.
FT DOMAIN 1436 1442 POLY-PRO.
FT DOMAIN 1982 1991 4 X 57 AA TYPE A REPEATS.
FT DOMAIN 1982 1991 1-1.
FT DOMAIN 41 95 1-2.
FT REPEAT 98 154 1-2.
FT REPEAT 281 337 1-3.
FT REPEAT 1740 1795 1-4.
FT REPEAT 337 549 2 X TYPE B REPEATS.
FT REPEAT 337 418 2-1.
FT REPEAT 476 549 2-2.
FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.
FT REPEAT 1899 1948 3-1.
FT REPEAT 1965 2014 3-2.
FT REPEAT 2040 2089 3-3.
FT REPEAT 57 57 R -> A (IN REF. 2).
FT CONFLICT 109 109 Q -> S (IN REF. 2).
FT CONFLICT 414 414 P -> PPHRGPGAGNGWPP (IN REF. 2).
FT CONFLICT 532 532 T -> K (IN REF. 2).
FT CONFLICT 682 682 Q -> K (IN REF. 2).
FT CONFLICT 730 730 E -> D (IN REF. 2).
FT CONFLICT 750 750 L -> T (IN REF. 2).
FT CONFLICT 834 834 A -> T (IN REF. 2).
FT CONFLICT 1035 1035 G -> A (IN REF. 2).
FT CONFLICT 1068 1068 M -> L (IN REF. 2).
FT CONFLICT 1285 1285 P -> R (IN REF. 2).
FT CONFLICT 1400 1400 G -> A (IN REF. 2).
FT CONFLICT 1611 1611 T -> S (IN REF. 2).
FT CONFLICT 1729 1729 G -> A (IN REF. 2).
FT CONFLICT 2142 AA; 227840 MW; 32DDF16B9B52420A CRC64;
SQ QUERY MATCH 27.1%; Score 52; DB 1; Length 2142;
Best Local Similarity 44.8%; Pred. No. 2.7e+02;
Matches 13; Conservative 2; Mismatches 6; Indels 8; Gaps 2;

QY 3 GPTLR-----QWLAARAGGGGGGIEGP 26
DB 196 GPSLRPQNSTTW---RDGGGRGPDELEGP 221

RESULT 19
NYLB_FLASK STANDARD; PRT; 392 AA.
ID NYLB_FLASK
AC P07061;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 6-aminohexanoate-dimer hydrolase (EC 3.5.1.46) (Nylon oligomers
degrading enzyme EII).
GN NYLB.
OS Flavobacterium sp. (strain KI72).
OC Bacteria; CFB group; Flavobacteria; Flavobacteriaceae; Flavobacterium.
OC NCBI_TaxID=261;
OX NCBI_TaxID=261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84068129; PubMed=6646204;
RA Okada H., Negoro S., Kimura H., Nakamura S.;
RT "Evolutionary adaptation of plasmid-encoded enzymes for degrading
nylon oligomers.";
RL Nature 306:203-206(1983).

[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=85054785; PubMed=6389532;
RA Negoro S., Nakamura S., Kimura H., Fujiyama K., Zhang Y.Z.,
RA Kanaki N., Okada H.;
RT "Construction of hybrid genes of 6-aminohexanoic acid-oligomer
hydrolase and its analogous enzyme. Estimation of the intramolecular
regions important for the enzyme evolution.";
J. Biol. Chem. 259:13648-13651(1984).
RN RP ACTIVE SITE.
RX MEDLINE=90076168; PubMed=2512123;
RA Negoro S., Mitamura T., Oka K., Kanagawa K., Okada H.;
RT "Determination of the active-site serine of 6-aminohexanoate-dimer
hydrolase.";
Eur. J. Biochem. 185:521-524(1989).
CC -|- CATALYTIC ACTIVITY: N-(6-aminohexanoyl)-6-aminohexanoate + H(2)O =
2 6-aminohexanoate.
CC -|- PATHWAY: SECOND STEP IN THE DEGRADATION OF 6-AMINOHEXANOIC ACID
CYCLIC DIMER, A BY-PRODUCT OF NYLON MANUFACTURE.
CC -|- MISCELLANEOUS: THE EII ENZYME IS 100 TIMES MORE ACTIVE TOWARD THE
SUBSTRATE THAN THE EII' ENZYME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X00046; CAA24927.1; -.
CC EMBL; D26094; BAA05087.1; -.
DR PIR; A22644; A22644.
DR PIR; A29516; A29516.
DR PIR; S06849; S06849.
KW Nylon degradation; Hydrolase; Plasmid.
FT ACT SITE 112 112
SQ SEQUENCE 392 AA; 42693 MW; 9CF34C393C3E53D9 CRC64;

Query Match 26.8%; Score 51.5; DB 1; Length 392;
Best Local Similarity 45.7%; Pred. No. 65;
Matches 16; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

QY 2 EGPTRLQMLAARAGGGGSG-GIEGPTLRQWLAAR 35
DB 323 EGSYRQWWTCTNERNVSGIGIHQNL--WLDPR 355

RESULT 20
LACL_NEUCR STANDARD; PRT; 619 AA.
ID LACL_NEUCR
AC P06811;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
DE (Urishiol oxidase) (Laccase allele OR).
GN LACC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88087214; PubMed=2961749;
RA German U.A., Mueller G., Hunziker P.E., Lerch K.;
RT "Characterization of two allelic forms of Neurospora crassa laccase.
Amino- and carboxyl-terminal processing of a precursor.";
J. Biol. Chem. 263:885-896(1988).
RL [2]
RN SEQUENCE OF 379-619 FROM N.A.
RP SEQUENCE OF 87067412; PubMed=2947240;
RX MEDLINE=87067412; PubMed=2947240;

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X56451; CAA39832.1; -;
 DR PIR; S15018; S15018.
 DR HSSP; P11831; 1SR5.
 DR TRANSFAC; T00763; -;
 DR InterPro; IPR002100; MADS-box.
 DR Pfam; PF00319; SRE-TE; 1.
 DR PRINTS; PR00404; MADSDOMAIN.
 DR SMART; SM00432; MADS; 1.
 DR PROSITE; PS00350; MADS_BOX.1; 1.
 DR PROSITE; PS00066; MADS_BOX.2; 1.
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
 KW Phosphorylation.
 FT DOMAIN 98 152 MADS.
 SQ SEQUENCE 448 AA; 46115 MW; B3CDCA7E0D97C23B CRC64;

Query Match 27.1%; Score 52; DB 1; Length 448;
 Best Local Similarity 64.7%; Pred. No. 65;
 Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 10 LAARAGGNGSGGIEGP 26
 II III III III III
 DB 18 LARRAGNGAGCGPIRGP 34

RESULT 17
 ID FXD2_HUMAN STANDARD; PRT; 497 AA.
 AC O60548;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Headbox protein D2 (Forkhead-related protein FKHL17) (Forkhead-
 DE related transcription factor 9) (FREAC-9).
 GN FOXD2 OR FKHL17 OR FREAC9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98066765; PubMed=9403061;
 RA Ernstsson S., Betz R., Lagercrantz S., Larsson C., Ericksson S.,
 RA Cedberg A., Carlsson P., Enerbaeck S.;
 RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
 RT expressed human forkhead gene that maps to chromosome 1p32-p34.1;
 RL Genomics 46:78-85(1997).
 RN [2]
 RP REVISIONS.
 RA Enerbaeck S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
 CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF042832; AAC15421.1; -;
 DR HSSP; Q63245; 2HPH.

DR TRANSFAC; T02485; -;
 DR MIM; 602211; -;
 DR InterPro; IPR001766; Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD.1; 1.
 DR PROSITE; PS00658; FORK_HEAD.2; 1.
 DR PROSITE; PS00039; FORK_HEAD.3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation.
 FT DOMAIN 90 94
 FT DOMAIN 101 104 POLY-ALA.
 FT DNA_BIND 126 217 FORK-HEAD.
 FT DOMAIN 247 250 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 398 409 POLY-GLY.
 FT DOMAIN 421 426 POLY-GLY.
 FT DOMAIN 442 445 POLY-ALA.
 SQ SEQUENCE 497 AA; 49007 MW; EAAF498D216BE019 CRC64;

Query Match 27.1%; Score 52; DB 1; Length 497;
 Best Local Similarity 61.9%; Pred. No. 71;
 Matches 13; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 4 PT--LRQWLARAGGGNGSGG 22
 II III III III III
 DB 385 PTALLRQGLKTDAGGGAGGG 405

RESULT 18
 ID BAT2_HUMAN STANDARD; PRT; 2142 AA.
 AC P48634;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE Large proline-rich protein BAT2 (HLA-B-associated transcript 2).
 DE GN BAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=T-cell;
 RA Banerji J., Sands J., Strominger J.L., Spies T.;
 RT "A gene pair from the human major histocompatibility complex encodes
 RT large proline-rich proteins with multiple repeated motifs and a
 RT single ubiquitin-like domain."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
 RN [2]
 RP SEQUENCE OF 1-1860 FROM N.A.
 RX MEDLINE=93272029; PubMed=8499947;
 RA Iris F.J.M., Bougueleret L., Prieur S., Caterina D., Primas G.,
 RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,
 RA Cohen D.;
 RT "Dense Alu clustering and a potential new member of the NF kappa B
 RT family within a 90 kilobase HLA class III segment."
 RL Nat. Genet. 3:137-145(1993).
 CC -!- FUNCTION: UNKNOWN.
 CC -!- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M33509; AAA35585.1; -;
 DR EMBL; M33518; AAA35586.1; -;


```

FT PROPEP      25      80
FT CHAIN       81     210
FT DISULFID    97     170
FT DISULFID   141     199
FT DISULFID   158     201
FT CARBOHYD    76      76
SQ SEQUENCE   210 AA; 22426 MW; DBC6A30195E139AD CRC64;

Query Match
Best Local Similarity 27.3%; Score 52.5; DB 1; Length 210;
Matches 13; Conservative 3; Mismatches 15; Indels 9; Gaps 1;

QY 3 GPTLRQWL-----AARAGGGGSGGIEGPTLRQWLA 33
DB 129 GSPLRQYFFETRCADNAEAGGPGAGGGCGRGVDRRHWS 168

RESULT 15
SIX3_HUMAN
ID SIX3_HUMAN STANDARD; PRT; 332 AA.
AC O95343;
DT 15-JUL-1999 (Rel. 38, Created)
DT 01-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Homeobox protein SIX3 (Sine oculis homeobox homolog 3).
GN SIX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99107815; PubMed=9899003;
RA Grandino B., Gallardo M.E., Lopez-Rios J., Sanz R., Ramos C.,
RA Ayuso C., Bovolenta P., Rodriguez de Cordoba S.;
RT "Genomic cloning, structure, expression pattern, and chromosomal
RT location of the human SIX3 gene.";
RL Genomics 55:100-105(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377859; PubMed=10454822;
RA Leppert G.S., Yang J.-M., Sundin O.H.;
RT "Sequence and location of SIX3, a homeobox gene expressed in the human
RT eye.";
RL Ophthalmic Genet. 20:1-15(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Clark B.J., Hanson I.M., Brown A.G., Ferrier R.K., Prosser J.,
RA van Heyningen V.;
RT "SIX3, a member of the Sine oculis/Six family of transcription
RT factors, is expressed in the developing and adult human eye.";
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
RN [4]
RP VARIANTS HPE2 VAL-226; ALA-250 AND PRO-257.
RX MEDLINE=99295940; PubMed=10369266;
RA Wallis D.E., Roessler E., Hehr U., Nanni L., Wiltshire T.,
RA Richieri-Costa A., Gillesen-kaesbach G., Zackai E.H., Rommens J.,
RA Muenke M.;
RT "Mutations in the homeodomain of the human SIX3 gene cause
RT holoprosencephaly.";
RL Nat. Genet. 22:196-198(1999).
CC -!- FUNCTION: MAY BE INVOLVED IN VISUAL SYSTEM DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: DEFECTS IN SIX3 ARE THE CAUSE OF HOLOPROSENCEPHALY TYPE 2
CC (HPE2); A COMMON, SEVERE MALFORMATION OF THE BRAIN THAT INVOLVES
CC SEPARATION OF THE CENTRAL NERVOUS SYSTEM INTO LEFT AND RIGHT
CC HALVES.
CC -!- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEODOMAIN
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF092047; AAD11939.1; -
CC EMBL: AF049339; AAD15753.1; -
CC EMBL: AF083891; AAD51091.1; -
CC EMBL: AJ012611; CAB42539.1; -
CC HSSP: P40424; 1B72.
CC MIM: 603714; -
CC MIM: 157170; -
CC InterPro: IPR000047; HTH_repressr.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC PRINTS: PR00031; HTHREPRESSR.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEBOX_1; FALSE_NEG.
CC Developmental protein; Homeobox; DNA-binding; Nuclear protein;
CC Disease mutation; Holoprosencephaly.
CC DOMAIN 33 69 POLY-GLY.
CC DNA_BIND 206 265 HOMEBOX.
CC DOMAIN 263 266 POLY-ALA.
CC VARIANT 226 226 L->V (IN HPE2).
CC /FTID=VAR_003771.
CC VARIANT 250 250 V->A (IN HPE2).
CC /FTID=VAR_003772.
CC VARIANT 257 257 R->P (IN HPE2).
CC /FTID=VAR_003773.
CC SQ SEQUENCE 332 AA; 35486 MW; 21EA07F6A2DD978F CRC64;

Query Match 27.1%; Score 52; DB 1; Length 332;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 ROWLAARAGGGNGSGG 22
DB 25 RSILLASSGGNGCAGG 40

RESULT 16
SRF_XENLA
ID SRF_XENLA STANDARD; PRT; 448 AA.
AC P23790;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Serum response factor (SRF).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91184140; PubMed=2009862;
RA Mohun T.J., Chambers A.E., Towers N., Taylor M.V.;
RT "Expression of genes encoding the transcription factor SRF during
RT early development of Xenopus laevis: identification of a CARG
RT box-binding activity as SRF.";
RL EMO J. 10:933-940(1991).
CC -!- FUNCTION: SRF IS A TRANSCRIPTION FACTOR THAT BINDS TO THE SERUM
CC RESPONSE ELEMENT (SRE), A SHORT SEQUENCE OF DYAD SYMMETRY LOCATED
CC 300 BP TO THE 5' OF THE SITE OF TRANSCRIPTION INITIATION OF SOME
CC GENES.
CC -!- SUBUNIT: BINDS DNA AS A MULTIMER, PROBABLY A DIMER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: PHOSPHORYLATED (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS.

```

DR PROSITE; PS01036; HSP70_3; 1.
KW ATP-binding; Chaperone; Heat shock; Multigene family.
SQ SEQUENCE 631 AA; 69199 MW; 01ACA20600C9322F CRC64;

Query Match 27.6%; Score 53; DB 1; Length 631;
Best Local Similarity 37.0%; Pred. No. 69;
Matches 10; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 4 PTLROWLAARAGGNGSGGIEGPTLRQ 30
DB 603 PIISKLYGGPGGGSGSGSGPTTEE 629

RESULT 13
ID Z219 HUMAN STANDARD; PRT; 722 AA.
AC Q9P2Y4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 219.
GN ZNF219.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP Tissue-Tetis;
RX MEDLINE=20277481; PubMed=10819330;
RA Sakai T., Toyoda A., Hashimoto K., Maeda H.;
RT "Isolation and characterization of a novel zinc finger gene, ZNF219,
and mapping to the human chromosome 14q11 region.";
RL DNA Res. 7:137-141(2000).
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL; AB015427; BAA90526.1; -
CC MIM; 605036; -
CC InterPro; IPR000822; Znf-C2H2.
CC Pfam; PF00096; zf-C2H2; 9.
CC SMART; SM00355; ZnfC2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 57 520 ZINC FINGERS.
FT ZN_FING 57 79 C2H2-TYPE.
FT ZN_FING 85 107 C2H2-TYPE.
FT ZN_FING 163 186 C2H2-TYPE.
FT ZN_FING 274 296 C2H2-TYPE.
FT ZN_FING 302 324 C2H2-TYPE.
FT ZN_FING 498 520 C2H2-TYPE.
SQ SEQUENCE 722 AA; 76876 MW; B19DA77B148B45B CRC64;

Query Match 27.6%; Score 53; DB 1; Length 722;
Best Local Similarity 40.0%; Pred. No. 79;
Matches 18; Conservative 3; Mismatches 14; Indels 10; Gaps 2;

QY 2 EPTLROWLAARAGG--NGSGGIEG-----TLROWLAARA 36
DB 374 EPPSLGLYLSLRAGRGNGEGEPGGRSGGFRPLSSALPARA 418

RESULT 14
NT5_HUMAN
ID NT5_HUMAN STANDARD; PRT; 210 AA.
AC P34130;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-5 precursor (NT-5) (Neutrophin factor 5) (Neurotrophin-4)
DE (NT-4) (Neutrophin factor 4).
GN NT5 OR NTF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP Tissue=Prostate;
RX MEDLINE=92212967; PubMed=1313578;
RA Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
RA Belluscio L., Le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
RA Yancopoulos G.D.;
RT "Mammalian neurotrophin-4: structure, chromosomal localization,
RT tissue distribution, and receptor specificity.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=92075279; PubMed=1742028;
RX Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
RA Rosenthal A.;
RT "Neurotrophin-5: a novel neurotrophic factor that activates trk and
RT trkB.";
RL Neuron 7:857-866(1991).
[3]
RN X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
RX MEDLINE=20095835; PubMed=10631974;
RA Robinson R.C., Radziejewski C., Spraggon G., Greenwald J.,
RA Kostura M.R., Bartnick L.D., Stuart D.I., Choe S., Jones E.Y.;
RT "The structures of the neurotrophin 4 homodimer and the brain-derived
RT neurotrophic factor/neurotrophin 4 heterodimer reveal a common Trk-
RT binding site.";
RL Protein Sci. 8:2589-2597(1999).
CC -!- FUNCTION: TARGET-DERIVED SURVIVAL FACTOR FOR PERIPHERAL SENSORY
CC SYMPATHETIC NEURONS.
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN PROSTATE. LOWER LEVELS
CC IN THYMUS, PLACENTA, AND SKELETAL MUSCLE. EXPRESSED IN EMBRYONIC
CC AND ADULT TISSUES.
CC -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL; M86528; AAA60154.1; -
CC PIR; JH0503; JH0503.
CC PIR; A42687; A42687.
CC PDB; 1B8M; 09-FEB-99.
CC PDB; 1B98; 26-FEB-99.
CC MIM; 162662; -
CC InterPro; IPR002072; NGF.
CC Pfam; PF00243; NGF; 1.
CC PRINTS; PR00268; NGF.
CC ProDom; PD002052; NGF; 1.
CC SMART; SM00140; NGF; 1.
CC PROSITE; PS00248; NGF_1; 1.
CC PROSITE; PS02070; NGF_2; 1.
KW Growth factor; Signal; 3D-structure.
FT SIGNAL 1 24 POTENTIAL.

AC Q07700;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA polymerase I (EC 2.7.7.7) (POL I).
 GN POLA OR RV1629 OR MT1665 OR MTCV01B2.21.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RP SEQUENCE FROM N.A.
 RN [1]
 RC STRAIN=H37RV;
 RX MEDLINE=94124016; PubMed=8294019;
 RA Mirahi V., Huberts P., Davies S.S., Dudding L.R.;
 RT "A PCR method for the sequence analysis of the gyrA, polA and rnhA
 RL gene segments from mycobacteria";
 RL Gene 136:287-290(1993).
 RN [2]
 RC STRAIN=H37RV;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RL complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
 CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
 CC + (DNA)(N).
 CC -1- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; L11920; AAB46393.1; --
 DR EMBL; Z95554; CAB08882.1; --
 DR EMBL; AE007030; AAK45935.1; --
 DR HSSP; P19821; 1BGX.
 DR TIGR; MT1665; --
 DR TubercuList; Rv1629; --
 DR InterPro; IPR002562; 3_5_exonuclease.
 DR InterPro; IPR002421; 5_3_exonuclease.
 DR InterPro; IPR002298; DNA_poli.
 DR InterPro; IPR001098; DNA_pol_A.
 DR InterPro; IPR000513; Exo_N.I.
 DR InterPro; IPR003583; HHH_1.
 DR InterPro; IPR003584; HHH_2.
 DR Pfam; PF01367; 5_3_exonuclease; 1.

DR Pfam; PF02739; 5_3_exonuc_N; 1.
 DR Pfam; PF00476; DNA_pol_A; 1.
 DR PRINTS; PR00868; DNAPOLI.
 DR SMART; SM00474; 3SEXOC; 1.
 DR SMART; SM00475; 53EXOC; 1.
 DR SMART; SM00278; HHH1; 1.
 DR SMART; SM00279; HHH2; 1.
 DR SMART; SM00482; POLAC; 1.
 DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
 DR Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
 KW Hydrolase; Exonuclease; DNA-binding; Complete proteome.
 SQ SEQUENCE 904 AA; 98471 MW; 1C8E560FE5F74323 CRC64;

Query Match 28.1%; Score 54; DB 1; Length 904;
 Best Local Similarity 58.8%; Pred. No. 75;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 TLRQWLAARAGCGNGSG 21
 I:||||| ||| :|
 Db 325 TVRQWLAEHAGDGRRA 341

RESULT 10
 LSR2_MYCLE
 ID LSR2_MYCLE STANDARD; PRT; 112 AA.
 AC P24094;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE LSR2 protein precursor (15 kDa antigen) (A15).
 GN LSR2 OR M10234.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=92040086; PubMed=1840579;
 RA Sela S., Thole J.E., Ottenhoff T.H., Clark-Curtiss J.E.;
 RT "Identification of Mycobacterium leprae antigens from a cosmid
 RT library: characterization of a 15-kilodalton antigen that is
 RT recognized by both the humoral and cellular immune systems in leprosy
 RT patients";
 RL Infect. Immun. 59:4117-4124(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 RN [3]
 RP SEQUENCE OF 24-115 FROM N.A.
 RC MEDLINE=91126054; PubMed=1992456;
 RA Laal S., Sharma Y.D., Prasad H.K., Murtaza A., Singh S., Tangri S.,
 RA Misra R.S., Nath I.;
 RT "Recombinant fusion protein identified by lepromatous sera mimics
 RT native Mycobacterium leprae in T-cell responses across the leprosy
 RT spectrum";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:1054-1058(1991).
 CC -1- FUNCTION: DOMINANT T-CELL ANTIGEN AND STIMULATES
 CC LYMPHOPROLIFERATION.
 CC -1- DISEASE: MOST PROBABLY CAUSES THE LYMPHOPROLIFERATIVE RESPONSES
 CC OCCURRING IN LEPROSY.

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92212967; PubMed=1313578;
 RA IP N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
 RA Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
 RA Yancopoulos G.D.;
 RT "Mammalian neurotrophin-4: structure, chromosomal localization,
 RT tissue distribution, and receptor specificity";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92075279; PubMed=1742028;
 RA Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
 RA Rosenthal A.;
 RT "Neurotrophin-5: a novel neurotrophic factor that activates trk and
 RT trkB";
 RL Neuron 7:857-866(1991).
 CC -!- FUNCTION: COULD SERVE AS A TARGET-DERIVED TROPHIC FACTOR FOR
 CC SENSORY AND SYMPATHETIC NEURONS.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, MUSCLE, OVARY, BRAIN,
 CC HEART, STOMACH AND KIDNEY. EXPRESSED IN BOTH EMBRYO AND ADULT
 CC TISSUES.
 CC -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M86742; AAA41728.1;
 DR EMBL; S69323; AAB20548.1;
 DR PIR; JH0504; JH0504
 DR PIR; B42687; B42687.
 DR HSP; P34130; I98M.
 DR InterPro: IPR002072; NGF.
 DR Pfam: PF00243; NGF; 1.
 DR PRINTS; PR00268; NGF.
 DR ProDom; PD002052; NGF; 1.
 DR SMART; SM00140; NGF; 1.
 DR PROSITE; PS00248; NGF; 1;
 DR PROSITE; PS0270; NGF; 2; 1.
 KW Growth factor; Signal.
 FT SIGNAL 1 21
 FT PROPEP 22 79
 FT CHAIN 80 209
 FT DISULFID 96 169
 FT DISULFID 140 198
 FT DISULFID 157 200
 FT CARBOHYD 75 75
 FT CONFLICT 177 177
 SQ SEQUENCE 209 AA; 22332 MW; DF5112C05C5D5B85 CRC64;
 Query Match 28.9%; Score 55.5; DB 1; Length 209;
 Best Local Similarity 35.0%; Pred. No. 13;
 Matches 14; Conservative 2; Mismatches 15; Indels 9; Gaps 1;
 QY 3 GPTLRQL-----AARAGGNGSGGEGPTLRQLA 33
 Db 128 GSPLRQYFFETRCRAESAGGGPGVGGGGRVDRRHLS 167
 RESULT 8
 ID TBX2_HUMAN
 AC Q13207; O16424; STANDARD; PRT; 702 AA.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE T-box transcription factor TBX2 (T-box protein 2).
 GN TBX2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RX TISSUE=Fetal kidney;
 RX MEDLINE=96015055; PubMed=8530034;
 RA Campbell C., Goodrich K., Casey G., Beatty B.;
 RT "Cloning and mapping of a human gene (TBX2) sharing a highly conserved
 RT protein motif with the Drosophila omb gene";
 RL Genomics 28:255-260(1995).
 [2]
 RP SEQUENCE OF 152-245 FROM N.A.
 RX TISSUE=Fetal kidney;
 RX MEDLINE=96169568; PubMed=8597636;
 RA Law D.J., Gebuhr T., Garvey N., Agulnik S.I., Silver L.M.;
 RT "Identification, characterization, and localization to chromosome
 RT 17q21-22 of the human TBX2 homolog, member of a conserved
 RL mammalian gene family";
 RL Mamm. Genome 6:793-797(1995).
 CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES
 CC REQUIRED FOR MESODERM DIFFERENTIATION. PROBABLY PLAYS A ROLE IN
 CC LIMB PATTERN FORMATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ADULT IN KIDNEY, LUNG,
 CC AND PLACENTA. WEAK EXPRESSION IN HEART AND OVARY.
 CC -!- SIMILARITY: CONTAINS 1 T-BOX DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U28049; AAA73861.1;
 DR EMBL; S81264; AAB36216.1;
 DR HSP; P24781; 1XBR.
 DR MIM; 600747;
 DR InterPro: IPR001699; T-box.
 DR Pfam; PF00907; T-box; 1.
 DR PRINTS; PR00937; TBOX.
 DR SMART; SM00425; TBOX; 1.
 DR PROSITE; PS01283; TBOX_1; 1.
 DR PROSITE; PS01264; TBOX_2; 1.
 DR PROSITE; PS02522; TBOX_3; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein;
 FT Developmental protein.
 FT DOMAIN 48 63
 FT DNA_BIND 104 277
 FT T-BOX 107 517
 FT DOMAIN 507 517
 FT POLY-GLY. 571 579
 FT DOMAIN 585 593
 FT POLY-ALA. 593 593
 FT CONFLICT 155 155
 FT CONFLICT 165 168
 FT Y -> D (IN REF. 2).
 FT AGKA -> TDKT (IN REF. 2).
 SQ SEQUENCE 702 AA; 74194 MW; C6477134C69D7C2C CRC64;
 Query Match 28.1%; Score 54; DB 1; Length 702;
 Best Local Similarity 70.6%; Pred. No. 60;
 Matches 12; Conservative 1; Mismatches 2; Indels 2; Gaps 1;
 QY 10 LAARAGGNGSGGIEGP 26
 Db 502 LASVAGGNGGGG--GP 516
 RESULT 9
 ID DPOL_MYCTU
 ID DPOL_MYCTU STANDARD; PRT; 904 AA.

Wed Oct 9 10:30:16 2002

Matches 13; Conservative 6; Mismatches 7; Indels 13; Gaps 2;

QY 6 LROWLAA-----RAGGNGSGGI---EGPTLROW 31
 Db 48 LKQWVAGRRVGRGRRGSGRGAIDRQSAIREW 86

Query Match 30.5%; Score 58.5; DB 1; Length 105;
 Best Local Similarity 50.0%; Pred. No. 3.3;
 Matches 13; Conservative 3; Mismatches 7; Indels 3; Gaps 1;

QY 1 IEPTLRQWLAARAGGNGSGGIEGP 26
 Db 58 VEGP---QVGALEAGGPGAGGLEGP 80

RESULT 3
 LSR2_MYCTU STANDARD; PRT; 112 AA.

AC O06285; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE LSR2 protein precursor.
 GN LSR2 OR RV3597C OR MT3704 OR MTCY07H7B.25.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DOMINANT T-CELL ANTIGEN AND STIMULATES
 LYMPHOPROLIFERATION (BY SIMILARITY).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z95557; CAB08947.1; -;
 CC DR EMBL; AE007170; AAK48061.1; -;
 CC DR TIGR; MT3704; -;
 CC DR TubercuList; RV3597c; -;
 CC KW Antigen; Immune response; Signal; Complete proteome.
 CC FT SIGNAL 18 POTENTIAL.
 CC FT CHAIN 19 112 LSR2 PROTEIN.
 CC FT SEQUENCE 112 AA; 12098 MW; A4B32E478CBAC3E4 CRC64;
 CC SQ
 Query Match 30.5%; Score 58.5; DB 1; Length 112;
 Best Local Similarity 33.3%; Pred. No. 3.6;

RESULT 4
 SC02_HUMAN STANDARD; PRT; 266 AA.
 ID SC02_HUMAN
 AC O43819; Q9UR87;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE SC02 protein homolog, mitochondrial precursor.
 DE SC02.
 GN SC02.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Monocytes;
 RA Smink L.J., Burton J.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.
 RX MEDLINE=20014747; PubMed=10545952;
 RA Papadopoulos L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,
 Sadock J.E., Krishna S., Walker W., Selby J., Glerum D.M.,
 Van Coster R., Lyon G., Scalsis E., Lebel R., Kaplan P., Shanske S.,
 De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;
 RT "Fatal infantile cardioencephalomyopathy with COX deficiency and
 mutations in SC02, a COX assembly gene.";
 RL Nat. Genet. 23:333-337(1999).
 CC -!- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER
 TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE
 CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS
 CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND
 GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX
 ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX
 DEFICIENCIES

CC -----
 CC -!- SIMILARITY: BELONGS TO THE SC01/2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF177385; AAF05313.1; -;
 CC DR EMBL; AL021683; CAA16671.1; -;
 CC DR MIM; 604272; -;
 CC DR MIM; 604377; -;
 CC DR MIM; 220110; -;
 CC DR InterPro; IPR003782; SC01_SenC.
 CC DR Pfam; PF02630; SC01_SenC; 1.
 CC DR Mitochondrion; Transit peptide; Disease mutation; Polymorphism.
 CC KW TRANSIT 1 41 MITOCHONDRION (POTENTIAL).
 CC FT CHAIN 42 266 SC02 PROTEIN HOMOLOG.
 CC FT VARIANT 20 20 R -> P (IN DBSNP:140523).
 CC FT VARIANT 140 140 E -> K (IN FIC).
 CC FT VARIANT 225 225 S -> F (IN FIC).
 CC FT SEQUENCE 266 AA; 29810 MW; BC2F40E057329BF3 CRC64;
 CC SQ

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-32

Perfect score: 192

Sequence: 1 IEPTLRQWLAAAGGGSGGIEPTLRQWLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.5	30.5	105	1 INS_BOVIN	P01317 bos taurus
2	58.5	30.5	105	1 INS_SHEEP	P01318 ovis aries
3	58.5	30.5	112	1 LSR2_MYCTU	O06285 mycobacteri
4	58.5	30.5	266	1 SCO2_HUMAN	O43819 homo sapien
5	57.5	29.9	562	1 STR_ERPE	Q9yft9 aeropyrum p
6	56.5	29.4	969	1 PAC4_HUMAN	P29122 homo sapien
7	55.5	28.9	209	1 NT5_RAT	P34131 rattus norv
8	54	28.1	702	1 TBX2_HUMAN	Q13207 homo sapien
9	54	28.1	904	1 DP01_MYCTU	Q07700 mycobacteri
10	53.5	27.9	112	1 LSR2_MYCLE	P24094 mycobacteri
11	53	27.6	339	1 HAD9_MOUSE	P28357 mus musculu
12	53	27.6	631	1 HS73_BOVIN	P34933 bos taurus
13	53	27.6	722	1 L219_HUMAN	Q9p2y4 homo sapien
14	52.5	27.3	210	1 NT5_HUMAN	P34130 homo sapien
15	52	27.1	332	1 SLX3_HUMAN	Q95343 homo sapien
16	52	27.1	448	1 SRF_XENLA	P23790 xenopus lae
17	52	27.1	497	1 FXD2_HUMAN	O60548 homo sapien
18	52	27.1	2142	1 BAT2_HUMAN	P48634 homo sapien
19	51.5	26.8	392	1 NYLB_FLASK	P07061 flavobacter
20	51.5	26.8	619	1 LAC1_NEUCR	P06811 neurospora
21	51.5	26.8	619	1 LAC2_NEUCR	P10574 neurospora
22	51	26.6	201	1 YR21_TRSVR	P25245 tomato ring
23	51	26.6	482	1 G3B2_HUMAN	Q9un86 homo sapien
24	51	26.6	611	1 OM70_MOUSE	Q9czw5 mus musculu
25	51	26.6	824	1 ROU_HUMAN	Q00839 homo sapien
26	50.5	26.3	335	1 SPW1_RAT	P56225 rattus norv
27	50.5	26.3	620	1 Y870_METJA	O58280 methanococc
28	50.5	26.3	897	1 CAR1_DROME	Q24210 drosophila
29	50	26.0	165	1 LYCV_BPP21	P27359 bacterioph
30	50	26.0	165	1 LYCV_BPPA2	P10439 bacterioph
31	50	26.0	165	1 LYCV_ECOLI	P78285 escherichia
32	50	26.0	297	1 XERC_MYCLE	Q9cbu0 mycobacteri
33	50	26.0	307	1 CC36_CABEL	P34803 caenorhabdi

ALIGNMENTS

RESULT 1

ID	INS_BOVIN	STANDARD	PRT	105 AA
AC	P01317			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Insulin precursor.			
GN	INS.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88288209; PubMed=2456452;			
RA	D'Agostino J., Younes M.A., White J.W., Besch P.K., Field J.B.,			
RA	Frazier M.L.;			
RT	"Cloning and nucleotide sequence analysis of complementary			
RT	deoxyribonucleic acid for bovine preproinsulin.";			
RL	Mol. Endocrinol. 1:327-331(1987).			
RN	[2]			
RP	SEQUENCE OF 25-105.			
RX	MEDLINE=71166442; PubMed=4928892;			
RA	Nolan C., Margoliash E., Peterson J.D., Steiner D.F.;			
RT	"The structure of bovine proinsulin.";			
RL	J. Biol. Chem. 246:2780-2795(1971).			
RN	[3]			
RP	SEQUENCE OF 25-54.			
RA	Sanger F., Tuppy H.;			
RT	"The amino-acid sequence in the phenylalanyl chain of insulin. 2. The			
RT	investigation of peptides from enzymic hydrolysates.";			
RL	Biochem. J. 49:481-490(1951).			
RN	[4]			
RP	SEQUENCE OF 57-82.			
RX	MEDLINE=71116409; PubMed=5545080;			
RA	Steiner D.F., Cho S., Oyer P.E., Terris S., Peterson J.D.,			
RA	Rubenstein A.H.;			
RT	"Isolation and characterization of proinsulin C-peptide from bovine			
RT	pancreas.";			
RL	J. Biol. Chem. 246:1365-1374(1971).			
RN	[5]			
RP	SEQUENCE OF 57-82.			
RX	MEDLINE=71257721; PubMed=5105368;			
RA	Salokangas A., Smyth D.G., Markussen J., Sundby F.;			
RT	"Bovine proinsulin: amino acid sequence of the C-peptide isolated			
RT	from pancreas.";			
RL	Eur. J. Biochem. 20:183-189(1971).			
RN	[6]			
RP	SEQUENCE OF 85-105.			
RA	Sanger F., Thompson E.O.P.;			
RT	"The amino-acid sequence in the glycyl chain of insulin. 2. The			
RT	investigation of peptides from enzymic hydrolysates.";			
RL	Biochem. J. 53:366-374(1953).			
RN	[7]			

34	50	26.0	348	1 YZ17_AQUAE	O66408 aquifex aeo
35	50	26.0	478	1 MKR1_WACEU	Q9tt91 macropus eu
36	50	26.0	613	1 MOT8_HUMAN	P36021 homo sapien
37	50	26.0	694	1 FZD8_HUMAN	Q9H461 homo sapien
38	50	26.0	956	1 DP01_DEIRA	P52027 deinococcus
39	50	26.0	1461	1 IE18_PRVIF	P11675 pseudorabie
40	50	26.0	2073	1 FAS1_SCHPO	Q9uug0 s fatty aci
41	49.5	25.8	333	1 CBBR_XANFL	P25545 xanthobacte
42	49.5	25.8	385	1 NK3R_MOUSE	P47937 mus musculu
43	49.5	25.8	825	1 SE5_RAT	O63003 rattus norv
44	49.5	25.8	4303	1 PKD1_HUMAN	P98161 homo sapien
45	49	25.5	172	1 SSB_BACSU	P37455 bacillus su

```

C.Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000
C.Accession: T45448
R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z22967
A.Accession: T45448
A.Status: preliminary; translated from GB/EMBL/DDBJ
A.Molecule type: DNA
A.Residues: 1-452 <JAM>
A.Cross-references: EMBL:AL035500; PIDN:CAB36690.1
A.Experimental source: cosmid L373
C.Genetics:
A.Note: MLCB373.28
C.Keywords: hydrolase; serine proteinase
F:182,224,305/Active site: His, Asp, Ser #status predicted

Query Match          27.6%; Score 53; DB 2; Length 452;
Best Local Similarity 52.9%; Pred. No. 96;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 10 LAARAGGNGNGSGGIEGP 26
    | | | | | | | |
Db 111 LGTRVAGNGAGPVTGP 127

Search completed: October 9, 2002, 09:05:12
Job time : 10.0937 secs

```


Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:06 11 May 2015

C;Genetics:

A; CROSS-REFERENCES: 00:0000, 00:0000

A; CROSS-REFERENCES: 00:0000, 00:0000

C;Accession: G87033
 R;Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hrd
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
 A;Title: Massive gene decay in the leprosy bacillus.
 A;Reference number: A86909; MUID:21128732; PMID:11234002
 A;Accession: G87033
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-488 <STO>
 A;Cross-references: GB:AL450380; NID:g13093026; PIDN:CAC31378.1; GSPDB:GN00147
 C;Genetics:
 A;Gene: ML0997
 C;Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 29.4%; Score 56.5; DB 2; Length 488;
 Best Local Similarity 40.0%; Pred. No. 42;
 Matches 12; Conservative 2; Mismatches 9; Indels 7; Gaps 1;

QY 4 PTLRW-----LAARAGGNGSGGIRGP 26
 ||| | : |||| | : ||
 Db 189 PRLRGWGESMSRQVGGGAGGGVGLRGP 218

RESULT 11
 S72938
 hflX protein - Mycobacterium leprae
 A;Alternate names: B2235_C2_202 protein
 C;Species: Mycobacterium leprae
 C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
 A;Accession: S72938
 R;Smith, D.R.; Robison, K.
 submitted to the EMBL Data Library, November 1993
 A;Description: Mycobacterium leprae cosmid B2235.
 A;Reference number: S72587
 A;Accession: S72938
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-318 <SMI>
 A;Cross-references: EMBL:U00019; NID:g467079; PIDN:AAA17274.1; PID:g467091
 C;Genetics:
 A;Start codon: GTG
 C;Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 29.4%; Score 56.5; DB 2; Length 518;
 Best Local Similarity 40.0%; Pred. No. 44;
 Matches 12; Conservative 2; Mismatches 9; Indels 7; Gaps 1;

QY 4 PTLRW-----LAARAGGNGSGGIRGP 26
 ||| | : |||| | : ||
 Db 219 PRLRGWGESMSRQVGGGAGGGVGLRGP 248

RESULT 12
 JC2191
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form C - hum
 N;Alternate names: kexin-like protease isoform
 C;Species: Homo sapiens (man)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 20-Apr-2000
 A;Accession: JC2191
 R;Tsuiji, A.; Higashine, K.; Hine, C.; Mori, K.; Tamai, Y.; Nagamune, H.; Matsuda, Y.
 Biochem. Biophys. Res. Commun. 200, 943-950, 1994
 A;Title: Identification of novel cDNAs encoding human kexin-like protease, PACE4 isoform
 A;Reference number: JC2191; MUID:94235049
 A;Accession: JC2191
 A;Molecule type: mRNA
 A;Residues: 1-652 <TSU>
 C;Comment: This protein consists of a signal peptide, a propeptide, a subtilisin-like p
 C;Comment: This protein cleaves precursor proteins at dibasic amino acid residues.
 C;Genetics:
 A;Gene: GDB:PACE4

A;Cross-references: GDB:131390; OMIM:167405
 A;Map position: 15q26-15q26
 C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C;Keywords: alternative splicing; hydrolase; serine proteinase
 F;196-434/Domain: subtilisin homology <SBT>
 F;205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 29.4%; Score 56.5; DB 1; Length 652;
 Best Local Similarity 44.8%; Pred. No. 55;
 Matches 13; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 11 AARAGGNGSGGIEGPTLR-----QWL 32
 |||| |:|| | :||
 Db 24 AAGAGGAGGAGGAGGCGFRPLAPRPWRWL 52

RESULT 13
 JC5571
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form E-II
 C;Species: Homo sapiens (man)
 C;Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 20-Jun-2000
 C;Accession: JC5571
 R;Mori, K.; Kii, S.; Tsuji, A.; Nagahama, M.; Imamaki, A.; Hayashi, K.; Akamatsu, T.;
 J. Biochem. 121, 941-948, 1997
 A;Title: A novel human PACE4 isoform, PACE4E is an active processing protease contain
 A;Reference number: JC5570; MUID:97335942
 A;Accession: JC5571
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-962 <MOR>
 A;Cross-references: DDBJ:D87994; NID:g2330550; PIDN:BAA21792.1; PID:g2330551
 A;Experimental source: brain cerebellum
 C;Comment: This enzyme is a processing protease and responsible for processing of var
 ch it is retained intracellularly.
 C;Genetics:
 A;Gene: GDB:PACE4
 A;Cross-references: GDB:131390; OMIM:167405
 A;Map position: 15q26-15q26
 C;Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase
 F;1-62/Domain: signal sequence #status predicted <SIG>
 F;63-149/Domain: propeptide #status predicted <PRO>
 F;196-434/Domain: subtilisin homology <SBT>
 F;938-954/Domain: hydrophobic cluster #status predicted <HCL>
 F;205,246,347,420/Active site: Asp, His, Asn, Ser #status predicted
 F;259/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 29.4%; Score 56.5; DB 2; Length 962;
 Best Local Similarity 44.8%; Pred. No. 80;
 Matches 13; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 11 AARAGGNGSGGIEGPTLR-----QWL 32
 |||| |:|| | :||
 Db 24 AAGAGGAGGAGGAGGCGFRPLAPRPWRWL 52

RESULT 14
 A39490
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form A -
 N;Alternate names: kexin homology
 C;Species: Homo sapiens (man)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Mar-2000
 C;Accession: A39490
 R;Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landsberg, K.E.; Saltman, D.; Barr, P.J.
 DNA Cell Biol. 10, 757-769, 1991
 A;Title: Identification of a second human subtilisin-like protease gene in the fes/fp
 A;Reference number: A39490; MUID:92075167
 A;Accession: A39490
 A;Molecule type: mRNA
 A;Residues: 1-969 <KIE>
 A;Cross-references: GB:M80482; NID:gl89531; PIDN:AAA59998.1; PID:gl89532
 C;Genetics:
 A;Gene: GDB:PACE4

QY 3 GPTLROWLAARAGGGNGSGGIEG 25
| | :: | || | || |
Db 429 GSWLGRFLNRGGGGGGGGMG 451

RESULT 8
E87575
ABC transporter, ATP-binding protein CC2634 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87575
R: Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, R.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Koonin, J.; Elneliseva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: AB87249; MUID:21173698; PMID:11259647
A:Accession: E87575
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: GB:AE005673; NID:gl3424211; PIDN:AAK24601.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2634

Query Match 29.4%; Score 56.5; DB 2; Length 249;
Best Local Similarity 42.9%; Pred. No. 22;
Matches 15; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 IEPTLROWLAARAGGGNGSGGIEGPTLROWLAAR 35
| | | | :: | | | | : | | | | | | | |
Db 54 LEQPT-RGAVSRLAGKGETSVVFOAPTLPAPWLSAR 87

RESULT 9
B39490
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form B -
N:Alternate names: subtilisin homolog precursor, short splice form
C:Species: Homo sapiens (man)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 31-Mar-2000
C:Accession: B39490
R: Kiefer, M.C.; Tucker, J.B.; Joh, R.; Landsberg, K.E.; Saltman, D.; Barr, P.J. DNA Cell Biol. 10, 757-769, 1991
A:Title: Identification of a second human subtilisin-like protease gene in the fes/fp
A:Reference number: A39490; MUID:92075167
A:Accession: B39490
A:Molecule type: mRNA
A:Residues: 1-487 <IE>
A:Note: the lack of a domain necessary for correct folding and activity of other seri
C:Genetics:
A:Gene: GDB:PACE4
A:Cross-references: GDB:131390; OMIM:167405
A:Map position: 15q26-15q26
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: alternative splicing; hydrolase, serine proteinase
F:196-434/Domain: subtilisin homology <SBT>
F:205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 29.4%; Score 56.5; DB 2; Length 487;
Best Local Similarity 44.8%; Pred. No. 42;
Matches 13; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 11 AARAGGGNGSGGIEGPTLR-----QWL 32
| | | | | | | | | | | | | |
Db 24 AAGAGGAGGAGGAGGFGFRPLAPRPWRWL 52

RESULT 10
GB7033
probable ATP/GTP-binding protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May 2001

1

[REDACTED]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8.09368 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838c-32
Perfect score: 192
Sequence: 1 IEGPTLRQWLAAAGGNGSGIEGPTLRQWLAAAR 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	31.2	683	2 B71325	conserved hypothet
2	58.5	30.5	77	1 INSH	insulin precursor
3	58.5	30.5	105	1 IPBO	insulin precursor
4	58.5	30.5	112	2 F70954	probable lsr2 prot
5	57.5	29.9	562	2 F72771	probable lysyl-trn
6	57	29.7	200	2 E85047	hypothetical prote
7	57	29.7	500	2 T20961	hypothetical prote
8	56.5	29.4	249	2 E87575	ABC transporter, A
9	56.5	29.4	487	2 B39490	subtilisin-like pr
10	56.5	29.4	488	2 G87033	probable ATP/GTP-b
11	56.5	29.4	518	2 S72938	subtilisin-like pr
12	56.5	29.4	652	1 JC2191	subtilisin-like pr
13	56.5	29.4	962	2 JC5571	subtilisin-like pr
14	56.5	29.4	969	1 A39490	subtilisin-like pr
15	56.5	29.4	975	2 JC5570	CREB-binding prote
16	56	29.2	3190	2 T13828	neurotrophin-4 pre
17	55.5	28.9	209	2 B42687	transcription fact
18	55.5	28.9	514	2 A35658	splicing factor SF
19	55	28.6	303	2 S71185	phosphoglucosidase
20	55	28.6	545	2 D87259	glutathione S-tran
21	54.5	28.4	209	2 C87617	probable HflX - My
22	54.5	28.4	495	2 D70505	probable membrane
23	54	28.1	296	2 AG0147	LRG5 protein - Chl
24	54	28.1	640	2 T08179	T-box protein 2 -
25	54	28.1	702	2 G01840	probable polA prot
26	54	28.1	904	2 C70559	LSR2 T-cell antige
27	53.5	27.9	112	2 B43601	homeotic protein H
28	53	27.6	339	2 S20880	probable secreted
29	53	27.6	382	2 H86930	

ALIGNMENTS

RESULT 1

B71325

conserved hypothetical protein TP0421 - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999

C:Accession: B71325

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: B71325

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-683 <COL>

A:CROSS-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AC65409.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0421

Query Match 31.2%; Score 60; DB 2; Length 683;

Best Local Similarity 43.8%; Pred. No. 23;

Matches 14; Conservative 2; Mismatches 12; Indels 4; Gaps 1;

OY 4 PTLRQWLAAAGGNGSGIEGPTLRQWLAAAR 35

Db 74 PLLEWL----GNAYRSIGIEGAALHQAAR 101

| : : | | | | | | | | | | | | | | | |

| : : | | | | | | | | | | | | | | | |

RESULT 2

INSH

insulin precursor - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999

C:Accession: S16430; S16431

R:Brown, H.; Sanger, F.; Kitai, R.

Biochem. J. 60, 556-565, 1955

A:Title: The structure of pig and sheep insulins.

A:Reference number: A90344

A:Accession: S16430

A:Molecule type: protein

A:Residues: 1-30:57-77 <BRO>

R:Peterson, J.D.; Neirlich, S.; Oyer, P.E.; Steiner, D.F.

J. Biol. Chem. 247, 4866-4871, 1972

A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin

A:Reference number: A92111; MUID:72258016

A:Accession: S16431

A:Molecule type: protein

A:Residues: 31-56 <PET>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 38.0%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

RESULT 29

US-09-516-704-18
Sequence 18, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:

NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-516-704-18

Query Match 38.0%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 1 IEPTLRQWLAARA 14

RESULT 30

US-09-516-704-194
Sequence 194, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-09-516-704-194
Query Match 38.0%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

Search completed: October 9, 2002, 09:06:34
Job time : 5.98595 secs

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
US-09-244-298A-18

Query Match 38.0%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQLAARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 38.0%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQLAARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

us-09-422-838c-32.ra1

Wed Oct' 9 10:30:13 2002

```

;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18
    Query Match      38.0%; Score 73; DB 3; Length 16;
    Best Local Similarity 100.0%; Pred. No. 0.0063;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAA 14
    |||||||
Db 1 IEGPTLRQWLAA 14

RESULT 24
US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194
    Query Match      38.0%; Score 73; DB 3; Length 16;
    Best Local Similarity 100.0%; Pred. No. 0.0063;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAA 14
    |||||||
Db 2 IEGPTLRQWLAA 15

RESULT 25
US-08-973-225-220
; Sequence 220, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220
    Query Match      38.0%; Score 73; DB 3; Length 16;
    Best Local Similarity 100.0%; Pred. No. 0.0063;
    Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAA 14
    |||||||
Db 2 IEGPTLRQWLAA 15

RESULT 26
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Depreince, Randolph B.
```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-194

```

```

Query Match      38.0%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

```

```

RESULT 22
US-08-764-640-232
Sequence 232, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-232

```

```

Query Match      38.0%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

```

```

RESULT 23
US-08-973-225-18
Sequence 18, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Mattheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide

```

us-09-422-838c-32.ra

Wed Oct '9 10:30:13 2002

Db 1 IEGPTLROWLAARA 14

RESULT 19
US-09-516-704-185
; Sequence 185, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprence, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: peptide
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
; US-09-516-704-185

Query Match 38.0%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | | | |
Db 2 IEGPTLROWLAARA 15

RESULT 20
US-08-764-640-18
; Sequence 18, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprence, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244

US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprence, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244

QY 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | | | |
Db 1 IEGPTLROWLAARA 14

RESULT 21
US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprence, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244


```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-17

Query Match 38.0%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAA 14
Db 1 IEGPTLRQWLAA 14
|||||

RESULT 17
US-09-244-298A-185
; Sequence 185, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-185

Query Match 38.0%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAA 14
Db 2 IEGPTLRQWLAA 15
|||||

RESULT 18
US-09-516-704-17
; Sequence 17, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 38.0%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAA 14
|||||
```

RESULT 14

US-08-973-225-17
; Sequence 17, Application US/08973225A

; Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-08-973-225-17

Query Match 38.0%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 IEPTLRQWLAARA 14

|||||

Db

1 IEPTLRQWLAARA 14

RESULT 15

US-08-973-225-185

; Sequence 185, Application US/08973225A
; Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.

Wagstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-08-973-225-185

Query Match 38.0%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 IEPTLRQWLAARA 14

|||||

Db

2 IEPTLRQWLAARA 15

RESULT 16

US-09-244-298A-17

; Sequence 17, Application US/09244298A
; Patent No. 6121238

GENERAL INFORMATION:

APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-09-516-704-193

Query Match 38.0%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 12

US-08-764-640-17
Sequence 17, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 38.0%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 13

US-08-764-640-185
Sequence 185, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-185

Query Match 38.0%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPTLRQWLAAARA 15

us-09-422-838c-32.ra1

Wed Oct' 9 10:30:13 2002

```
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrublec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
;
; US-09-516-704-13
;
; Query Match 38.0%; Score 73; DB 4; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.0054;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQWLAAARA 14
;      |||||
; DB 1 IEGPTLRQWLAAARA 14
;
; RESULT 11
; US-09-516-704-193
; Sequence 193, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrublec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
; US-09-244-298A-193
;
; Query Match 38.0%; Score 73; DB 3; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.0054;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQWLAAARA 14
;      |||||
; DB 1 IEGPTLRQWLAAARA 14
;
; RESULT 10
; US-09-516-704-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
```

;; FILING DATE: 04-Dec-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hrubiec, Robert T.
;; REGISTRATION NUMBER: 36,392
;; REFERENCE/DOCKET NUMBER: PK3065USW
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-248-1000
;; INFORMATION FOR SEQ ID NO: 13:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 38.0%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 7
US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Matheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 38.0%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: DePrince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 38.0%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 9
US-09-244-298A-193
; Sequence 193, Application US/09244298A

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deptrince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 38.0%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 5
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deptrince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 38.0%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 6
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Mattheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A

NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
US-08-764-640-231

Query Match 39.8%; Score 76.5; DB 2; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0039;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 2 EGPTLROWLAARAGGNGSGGIEGPTLROWLA 33
:|||||: 1 :|||||: :
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 2
US-09-244-298A-231
; Sequence 231, Application US/09244298A
; Patent No. 6121238

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESS: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
US-09-244-298A-231

Query Match 39.8%; Score 76.5; DB 3; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0039;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 2 EGPTLROWLAARAGGNGSGGIEGPTLROWLA 33

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 3
US-09-516-704-231
; Sequence 231, Application US/09516704
; Patent No. 6251864

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESS: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
US-09-516-704-231

Query Match 39.8%; Score 76.5; DB 4; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0039;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 2 EGPTLROWLAARAGGNGSGGIEGPTLROWLA 33
:|||||: 1 :|||||: :
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 4
US-08-764-640-13
; Sequence 13, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:

GenCore version 5.1.3
Copyright.(c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 ; Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-32
Perfect score: 192
Sequence: 1 IEPTLRQWLAARAGGNGSGIEPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.5	39.8	25	2	US-08-764-640-231
2	76.5	39.8	25	3	US-09-244-298A-231
3	76.5	39.8	25	4	US-09-516-704-231
4	73	38.0	14	2	US-08-764-640-13
5	73	38.0	14	2	US-08-764-640-193
6	73	38.0	14	3	US-08-973-225-13
7	73	38.0	14	3	US-08-973-225-193
8	73	38.0	14	3	US-09-244-298A-13
9	73	38.0	14	3	US-09-244-298A-193
10	73	38.0	14	4	US-09-516-704-13
11	73	38.0	14	4	US-09-516-704-193
12	73	38.0	15	2	US-08-764-640-17
13	73	38.0	15	2	US-08-764-640-185
14	73	38.0	15	3	US-08-973-225-17
15	73	38.0	15	3	US-08-973-225-185
16	73	38.0	15	3	US-09-244-298A-17
17	73	38.0	15	3	US-09-244-298A-185
18	73	38.0	15	4	US-09-516-704-17
19	73	38.0	15	4	US-09-516-704-185
20	73	38.0	16	2	US-08-764-640-18
21	73	38.0	16	2	US-08-764-640-194
22	73	38.0	16	2	US-08-764-640-232
23	73	38.0	16	3	US-08-973-225-18
24	73	38.0	16	3	US-08-973-225-194
25	73	38.0	16	3	US-08-973-225-220
26	73	38.0	16	3	US-09-244-298A-18
27	73	38.0	16	3	US-09-244-298A-194

28	73	38.0	16	3	US-09-244-298A-232	Sequence 232, App
29	73	38.0	16	4	US-09-516-704-18	Sequence 18, Appl
30	73	38.0	16	4	US-09-516-704-194	Sequence 194, App
31	73	38.0	16	4	US-09-516-704-232	Sequence 232, App
32	69	35.9	14	2	US-08-764-640-195	Sequence 195, App
33	69	35.9	14	2	US-08-764-640-199	Sequence 199, App
34	69	35.9	14	3	US-08-973-225-195	Sequence 195, App
35	69	35.9	14	3	US-08-973-225-199	Sequence 199, App
36	69	35.9	14	3	US-09-244-298A-195	Sequence 195, App
37	69	35.9	14	3	US-09-244-298A-199	Sequence 199, App
38	69	35.9	14	4	US-09-516-704-195	Sequence 195, App
39	69	35.9	14	4	US-09-516-704-199	Sequence 199, App
40	69	35.9	15	2	US-08-764-640-196	Sequence 196, App
41	69	35.9	15	2	US-08-764-640-200	Sequence 200, App
42	69	35.9	15	2	US-08-764-640-209	Sequence 209, App
43	69	35.9	15	2	US-08-764-640-215	Sequence 215, App
44	69	35.9	15	3	US-08-973-225-196	Sequence 196, App
45	69	35.9	15	3	US-08-973-225-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-08-764-640-231
; Sequence 231, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Poddaturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

QY 1 IEGPTLRQWLAAAGGNGSGGIEGPTLRQWLAAARA 36
|||||
Db 1 IEGPTLRQWLAAARAGG----GGIEGPTLRQWLAAARA 32
|||||

Db 1 IEGPTLRQWLAAARAGGSGGGGGIEGPTLRQWLAAARA 36
|||||

Search completed: October 9, 2002, 08:58:58
Job time : 16.1874 secs

RESULT 30
AAB17298
ID AAB17298 standard; Peptide: 36 AA.
XX
AC AAB17298;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:354.
XX

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antitumoral; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX

OS Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

XX Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antitumoral, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can
XX be used for producing pharmaceutical compositions. The compositions are
XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
XX The use of an Fc domain (rather than a Fab domain) can provide a longer
XX half-life or incorporate functions such as Fc receptor binding, protein
XX A binding, complement fixation, and possibly placental transfer. AAA69443
XX to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
XX sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 81.2%; Score 156; DB 21; Length 36;
Best Local Similarity 88.9%; Pred. No. 2.9e-12;
Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGNGSGGIEGPTLRQWLAAARA 36

Query Match 85.9%; Score 165; DB 21; Length 34;
 Best Local Similarity 91.7%; Pred. NO. 2.2e-13;
 Matches 33; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1 IEGPTLQWLAAARAGGSGGIEGPTLQWLAAARA 36
 |||||
 Db 1 IEGPTLQWLAAARAGG--GGGIEGPTLQWLAAARA 34

RESULT 28
 AAB17290
 ID AAB17290 standard; Peptide; 33 AA.
 XX
 AC AAB17290;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:346.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 317; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 33 AA;

Query Match 85.7%; Score 164.5; DB 21; Length 33;

Best Local Similarity 91.7%; Pred. No. 2.5e-13;
 Matches 33; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 IEGPTLQWLAAARAGGSGGIEGPTLQWLAAARA 36
 |||||
 Db 1 IEGPTLQWLAAARAGG--GGIEGPTLQWLAAARA 33

RESULT 29
 AAB17289
 ID AAB17289 standard; Peptide; 32 AA.
 XX
 AC AAB17289;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:345.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 316; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 32 AA;

Query Match 82.3%; Score 158; DB 21; Length 32;

Best Local Similarity 88.9%; Pred. NO. 1.5e-12;
 Matches 32; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX

SQ Sequence 40 AA;

Query Match 88.5%; Score 170; DB 21; Length 40;
Best Local Similarity 85.0%; Pred. No. 6.5e-14;
Matches 34; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARAGGNN-----GSGGTEGPTLRQWLAAAR 36
|||||
Db 1 IEGPTLRQWLAAARAGGKBACGGGGGTEGPTLRQWLAAAR 40
|||||

RESULT 26

AAB17296
ID AAB17296 standard; Peptide; 42 AA.

XX AC AAB17296;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:352.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

XX PS Example 1; Page 319; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions can
CC be used for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 42 AA;

Query Match 88.0%; Score 169; DB 21; Length 42;
Best Local Similarity 81.0%; Pred. No. 9.1e-14;
Matches 34; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

QY 1 IEGPTLRQWLAAAR-----GGGNGSGGTEGPTLRQWLAAAR 36
|||||
Db 1 IEGPTLRQWLAAARAGGGGGGGGGGGTEGPTLRQWLAAAR 42
|||||

RESULT 27

AAB17291

ID AAB17291 standard; Peptide; 34 AA.

XX AC AAB17291;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:347.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

XX PS Example 1; Page 317; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions can
CC be used for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 34 AA;

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 39 AA;
 SQ Query Match 88.8%; Score 170.5; DB 21; Length 39;
 Best Local Similarity 87.2%; Pred. No. 5.5e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels 3; Gaps 1;
 QY 1 IEGPTLROWLAARAGG---NGSGGIEGPTLROWLAARA 36
 |||||
 Db 1 IEGPTLROWLAARAGGPEGGGGIEGPTLROWLAARA 39

RESULT 24
 AAB17305
 ID AAB17305 standard; Peptide: 39 AA.
 XX
 AC AAB17305;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:361.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX Synthetic.
 OS
 XX WO200024782-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 25-OCT-1999; 99WO-US25044.
 XX
 XX PR 23-OCT-1998; 98US-0105371.
 XX
 XX PR 22-OCT-1999; 99US-0428082.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX
 XX PS Example 1; Page 323; 608pp; English.
 XX
 XX CC The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 XX where P1, P2, P3, and P4 = are each independently sequences of
 XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
 XX independently linkers; and a, b, c, d, e, and f = are each independently
 XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
 XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 XX activities. DNAs, vectors and host cells from the present invention can
 XX be used for producing pharmaceutical compositions. The compositions are
 XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 XX The use of an Fc domain (rather than a Fab domain) can provide a longer
 XX half-life or incorporate functions such as Fc receptor binding, protein
 XX A binding, complement fixation, and possibly placental transfer. AAA69443

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 39 AA;
 SQ Query Match 88.8%; Score 170.5; DB 21; Length 39;
 Best Local Similarity 87.2%; Pred. No. 5.5e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels 3; Gaps 1;
 QY 1 IEGPTLROWLAARAGG---NGSGGIEGPTLROWLAARA 36
 |||||
 Db 1 IEGPTLROWLAARAGGPEGGGGIEGPTLROWLAARA 39

RESULT 25
 AAB17302
 ID AAB17302 standard; Peptide: 40 AA.
 XX
 AC AAB17302;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:358.
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX Synthetic.
 OS
 XX WO200024782-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 25-OCT-1999; 99WO-US25044.
 XX
 XX PR 23-OCT-1998; 98US-0105371.
 XX
 XX PR 22-OCT-1999; 99US-0428082.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX
 XX PS Example 1; Page 322; 608pp; English.
 XX
 XX CC The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 XX where P1, P2, P3, and P4 = are each independently sequences of
 XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
 XX independently linkers; and a, b, c, d, e, and f = are each independently
 XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
 XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 XX activities. DNAs, vectors and host cells from the present invention can
 XX be used for producing pharmaceutical compositions. The compositions are
 XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 XX The use of an Fc domain (rather than a Fab domain) can provide a longer
 XX half-life or incorporate functions such as Fc receptor binding, protein
 XX A binding, complement fixation, and possibly placental transfer. AAA69443

activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an FC domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as FC receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA694 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Sequence . 38 AA:

Very Match 89.1%; Score 171; DB 21; Length 38;
 1st Local Similarity 89.5%; Pred. No. 4.7e-14;
 Matches 34; Conservative 0; Mismatches 2; Indels

1 IEPTLRQWLAARA - GGGNGSGGIEGPTLRQWLAARA 36
1 IEPTLRQWLAARAGGGGGGGGIEGPTLRQWLAARA 38

RESULT 23

ID AAB17304 standard; Peptide; 39 AA.

AABI/304;

31-OCT-2000 (first entry)

TP0-mimetic peptide sequence SEQ ID NO:360.

Modified peptide: therapeutic agent; fusion; FC domain; cancer; autoimmune disease; cystostatic; antiasthmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist; MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; vascular endothelial growth factor; matrix metalloproteinase; asthma; thrombosis; pharmaceutical.

Synthetic.

N WO200024782-A2.

04-MAY-2000

25-OCT-1999; 99WO-US25044.

23-OCT-1998; 98US-0105371

500 0420002.

(AMGE-) AMGEN INC.

Feige U, Liu C, Cheetham J, Boone TC;

WPL; 2000-350702/30.

Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

Example 1; Page 323; 608pp; English

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)_a-F1-(X2)_b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -L1-, -(L1)_c-P1-, -(L1)_c-P1-(L2)_d-P2-, -(L1)_c-P1-(L2)_d-P2-(L3)_e-P3-, -(L1)_c-P1-(L2)_d-P2-(L3)_e-P3-(L4)_f-P4, where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX
 XX Sequence 36 AA;
 Query Match 93.2%; Score 179; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. NO. 4.7e-15;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRLWLAARAGGCGGGIEGPTLRLWLAARA 36
 |||||
 Db 1 IEPTLRLWLAARAGGCGGGIEGPTLRLWLAARA 36

RESULT 18

AA96524
 ID AAY96524 standard; peptide; 36 AA.

XX AC AAY96524;

XX AC
 XX 04-SEP-2000 (first entry)

XX DE Thrombopoietin mimetic peptide compound 5.

XX KW Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1

FT Peptide /note= "optionally linked to an Fc molecule"

FT 1..14 /label= TMP_1

FT Disulfide-bond 9..31 /note= "optional"

FT 15..22 /label= linker

FT 23..36 /label= TMP_2

XX WO200024770-A2.

XX PD 04-MAY-2000.

XX PF 22-OCT-1999; 99WO-US24834.

XX PR 23-OCT-1998; 98US-0105348.

XX PA (AMGE-) AMGEN INC.

XX PI Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX DR Thrombopoietic peptides which activate mpl receptors and increase the
 XX production of platelets or platelet precursors, useful for treatment of
 XX diseases which involve thrombocytopenia
 XX Claim 16; Page 62; 91pp; English.

XX CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP_1-(L_1)_TMP_2),
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

CC 10 to 14 residues in length comprising X2-X1_0, X2-X1_1, X2-X1_2,
 CC X2-X1_3, X2-X1_4, X1-X1_0, X1-X1_1, X1-X1_2, X1-X1_3, and
 CC X1-X1_4, X1-X1_5 = L, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
 CC X4 = F; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
 CC or E; X9 = W, Y or F; X1_0 = L, I, V, A, F, M, or K; X1_1 = A, I, V,
 CC L, F, S, T, K, H, or G; X1_2 = A, I, V, L, F, G, S, or Q; X1_3 = R, K,
 CC T, V, N, Q or G; X1_4 = A, I, V, L, F, T, R, E, or G; L1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 36 AA;

Query Match 93.2%; Score 179; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. NO. 4.7e-15;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRLWLAARAGGCGGGIEGPTLRLWLAARA 36
 |||||
 Db 1 IEPTLRLWLAARAGGCGGGIEGPTLRLWLAARA 36

RESULT 19

AAB16959

ID AAB16959 standard; Protein; 268 AA.

XX AC AAB16959;

XX AC
 XX 31-OCT-2000 (first entry)

XX DE FC-TMP-TMP protein sequence SEQ ID NO:8.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX DR N-PSDB; AAA69445.

XX PT Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX Example 2; Page 182-183; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC where p_1, p_2, p_3 , and p_4 = are each independently sequences of

cc independently linkers; and a, b, c, d, e, and f = are each independently

XX PS Example 2; Page 331; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 60 AA;

Query Match 94.8%; Score 182; DB 21; Length 60;

Best Local Similarity 94.4%; Pred. No. 3.5e-15;

Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGNGSGGIEGPTLRQWLAAARA 36

|||||

DB 2 IEPTLRQWLAAARAGGNGSGGIEGPTLRQWLAAARA 37

|||||

RESULT 14

AAB16960

ID AAB16960 standard; Protein; 269 AA.

XX AC AAB16960;

XX 31-OCT-2000 (first entry)

DT

DE TMP-TMP-Fc protein sequence SEQ ID NO:10.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX OS Homo sapiens.

OS Synthetic.

XX WO200024782-A2.

PN

PD 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

PF

XX 23-OCT-1998; 98US-0105371.

PR

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

PA

XX Feige U, Cheetham J, Boone TC;

PI

XX WPI; 2000-350702/30.

DR

XX N-PSDB; AAA69446.

DR

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases.

XX PS Example 2; Page 185-186; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 269 AA;

Query Match 94.8%; Score 182; DB 21; Length 269;

Best Local Similarity 94.4%; Pred. No. 1.7e-14;

Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGNGSGGIEGPTLRQWLAAARA 36

|||||

DB 2 IEPTLRQWLAAARAGGNGSGGIEGPTLRQWLAAARA 37

|||||

RESULT 15

AAY96531

ID AAY96531 standard; Protein; 269 AA.

XX AC AAY96531;

XX 04-SEP-2000 (first entry)

DT

DE Human IgG1 Fc TMP fusion protein.

XX KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;

KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;

KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX OS Homo sapiens.

XX WO200024770-A2.

PN

PD 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

PF

XX 23-OCT-1998; 98US-0105348.

PR

XX (AMGE-) AMGEN INC.

PA

XX Liu C, Feige U, Cheetham J;

PI

XX WPI; 2000-365108/31.

DR

XX N-PSDB; AAA29229.

XX Thrombopoietic peptides which activate mpl receptors and increase the

PT production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

XX Example 2A; Page 49-50; 91pp; English.

PS

XX A compound which binds to an mpl receptor comprising a thrombopoietin

CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],

CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

DR WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX Example 2; Page 327; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1-(L2)d-P2, -(L4)f-P4

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA9443

CC to AAA9526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 94.8%; Score 182; DB 21; Length 42;

Best Local Similarity 94.4%; Pred. No. 2.4e-15;

Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 36

DB 7 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 42

RESULT 12

AAAY96530

ID AAAY96530 standard; Protein; 42 AA.

XX

AC AAAY96530;

XX

DT 04-SEP-2000 (first entry)

XX

DE Thrombopoietin mimetic peptide.

XX

KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;

KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;

KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.

XX

OS Synthetic.

XX

PN WO200024770-A2.

XX

PD 04-MAY-2000.

XX

PF 22-OCT-1999; 99WO-US24834.

XX

PR 23-OCT-1998; 98US-0105348.

XX

PA (AMGE-) AMGEN INC.

XX

PI Liu C, Feige U, Cheetham J;

XX

DR WPI; 2000-365108/31.

DR N-PSDB; AAA29225.

XX

XX Thrombopoietic peptides which activate mpl receptors and increase the

PT production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

XX

PS Example 2A; Page 48; 91pp; English.

XX Overlapping oligonucleotides were used to construct a synthetic

CC gene encoding a thrombopoietin mimetic peptide (TMP), which

CC was then fused in-frame to the Fc region of the human IgG1 chain (see

CC AAY96529). A compound which binds to an mpl receptor comprising a TMP

CC dimer joined by a linker [TMP₁-(L₁)₁-TMP₂], is new. TMP₁ and TMP₂

CC are amino acid sequences varying from at least 10 to 14 residues in

CC length comprising X₂-X₁-0, X₂-X₁-1, X₂-X₁-2, X₂-X₁-3, X₂-X₁-4,

CC X₁-X₁-0, X₁-X₁-1, X₁-X₁-2, X₁-X₁-3, and X₁-X₁-4. X₁ = I, A,

CC V, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S;

CC X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N, or E; X₉ = W, Y or F;

CC X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, L, F, S, T, K, H, or E;

CC X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K, T, V, N, O or G; X₁₄ =

CC A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino

CC acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl

CC receptor which mediates the activity of endogenous thrombopoietin. The

CC TMPs are useful for increasing the production of platelets or platelet

CC precursors (e.g. megakaryocytes) in a mammal, which is useful for

CC treatment of diseases which involve thrombocytopenia, e.g. aplastic

CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus

CC associated ITP, and systemic lupus erythematosus.

XX

SQ Sequence 42 AA;

Query Match 94.8%; Score 182; DB 21; Length 42;

Best Local Similarity 94.4%; Pred. No. 2.4e-15;

Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 36

DB 7 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 42

RESULT 13

AAB17311

ID AAB17311 standard; Peptide; 60 AA.

XX

AC AAB17311;

XX

DT 31-OCT-2000 (first entry)

XX

DE Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.

XX

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200024782-A2.

XX

PD 04-MAY-2000.

XX

PF 25-OCT-1999; 99WO-US25044.

XX

PR 23-OCT-1998; 98US-0105371.

XX

PR 22-OCT-1999; 99US-0428082.

XX

PA (AMGE-) AMGEN INC.

XX

PI Feige U, Liu C, Cheetham J, Boone TC;

XX

DR WPI; 2000-350702/30.

XX

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX DR WPI; 2000-350702/30.
 XX PT Novel composition of matter comprising an Fc domain and
 XX PT pharmacologically active peptides, useful for treating cancer and
 XX PT autoimmune diseases -
 XX PS Disclosure; Page 313; 608pp; English.
 XX CC The present invention describes composition of matter (I) comprising an
 XX CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 XX CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 XX CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 XX CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 XX CC where P1, P2, P3, and P4 = are each independently sequences of
 XX CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 XX CC independently linkers; and a, b, c, d, e, and f = are each independently
 XX CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 XX CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 XX CC activities. DNAs, vectors and host cells from the present invention can
 XX CC be used for producing pharmaceutical compositions. The compositions are
 XX CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 XX CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 XX CC half-life or incorporate functions such as Fc receptor binding, protein
 XX CC A binding, complement fixation, and possibly placental transfer. AAA69443
 XX CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 XX CC sequences used in the exemplification of the present invention.
 XX SQ Sequence 42 AA;

Query Match 94.8%; Score 182; DB 21; Length 42;
 Best Local Similarity 94.4%; Pred. No. 2.4e-15;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 7 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAAARA 42
 |||||

RESULT 10
 AAB17282
 ID AAB17282 standard; Peptide; 42 AA.
 XX AC AAB17282;
 XX DT 31-OCT-2000 (first entry)
 XX DE TPO-mimetic peptide sequence SEQ ID NO:338.
 XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX OS Synthetic.
 XX PN WO200024782-A2.
 XX PD 04-MAY-2000.
 XX PF 25-OCT-1999; 99WO-US25044.
 XX PR 23-OCT-1998; 98US-0105371.
 XX PR 22-OCT-1999; 99US-0428082.
 XX PA (AMGE-) AMGEN INC.
 XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX PI WPI; 2000-350702/30.
 XX DR Novel composition of matter comprising an Fc domain and
 XX PT pharmacologically active peptides, useful for treating cancer and
 XX PT autoimmune diseases -
 XX PS Disclosure; Page 313; 608pp; English.
 XX CC The present invention describes composition of matter (I) comprising an
 XX CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 XX CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 XX CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 XX CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 XX CC where P1, P2, P3, and P4 = are each independently sequences of
 XX CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 XX CC independently linkers; and a, b, c, d, e, and f = are each independently
 XX CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 XX CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 XX CC activities. DNAs, vectors and host cells from the present invention can
 XX CC be used for producing pharmaceutical compositions. The compositions are
 XX CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 XX CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 XX CC half-life or incorporate functions such as Fc receptor binding, protein
 XX CC A binding, complement fixation, and possibly placental transfer. AAA69443
 XX CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 XX CC sequences used in the exemplification of the present invention.
 XX SQ Sequence 42 AA;

Query Match 94.8%; Score 182; DB 21; Length 42;
 Best Local Similarity 94.4%; Pred. No. 2.4e-15;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEPTTLQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||

RESULT 11
 AAB17308
 ID AAB17308 standard; Peptide; 42 AA.
 XX AC AAB17308;
 XX DT 31-OCT-2000 (first entry)
 XX DE Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.
 XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX OS Homo sapiens.
 XX OS Synthetic.
 XX PN WO200024782-A2.
 XX PD 04-MAY-2000.
 XX PF 25-OCT-1999; 99WO-US25044.
 XX PR 23-OCT-1998; 98US-0105371.
 XX PR 22-OCT-1999; 99US-0428082.
 XX PA (AMGE-) AMGEN INC.
 XX PI Feige U, Liu C, Cheetham J, Boone TC;

Query Match	94.8%	Score 182;	DB 21;	Length 36;
Best Local Similarity	94.4%	Pred. No. 2e-15;		
Matches 34;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAARA 36		
DB	1	IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAARA 36		
RESULT 8				
AAAY96528				
ID	AAAY96528	standard; peptide; 41 AA.		
XX				
AC	AAAY96528;			
XX				
DT	04-SEP-2000	(first entry)		
XX				
DE	Thrombopoietin mimetic peptide compound 9.			
XX				
KW	Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;			
KW	anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;			
KW	immunosuppressive; anti-inflammatory; linker.			
XX				
OS	Synthetic.			
XX				
FH	Key	Location/Qualifiers		
FT	Modified-site	1	/note= "optionally linked to an Fc molecule"	
FT	Peptide	6..19	/label= TMP_1	
FT	Peptide	20..27	/label= linker	
FT	Peptide	28..41	/label= TMP_2	
XX				
PN	WO200024770-A2.			
XX				
PD	04-MAY-2000.			
XX				
PF	22-OCT-1999;	99WO-US24834.		
XX				

XX PS Example 1; Page 321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 94.8%; Score 182; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 2e-15; Mismatches 0; Gaps 0;
Matches 34; Conservative 0; Indels 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGIEGPTLRQWLAAARA 36
|||||
Db 1 IEGPTLRQWLAAARAGGGGGIEGPTLRQWLAAARA 36

RESULT 6
AAY96523
ID AAY96523 standard; peptide; 36 AA.

XX AAY96523;
XX 04-SEP-2000 (first entry)
XX Thrombopoietin mimetic peptide compound 4.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX Synthetic.

Key	Location/Qualifiers
FT Modified-site 1	/note= "optionally linked to an Fc molecule"
FT Peptide 1..14	/label= TMP_1
FT Peptide 15..22	/label= linker
FT Modified-site 18	/note= "optionally modified by bromoacetyl or PEG"
FT Peptide 23..36	/label= TMP_2

XX WO200024770-A2.
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-US24834.
XX 23-OCT-1998; 98US-0105348.
XX (AMGE-) AMGEN INC.
XX Liu C, Feige U, Cheetham J;
XX PI

DR WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
XX production of platelets or platelet precursors, useful for treatment of
XX diseases which involve thrombocytopenia
XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
XX mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2],
XX is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
XX 10 to 14 residues in length comprising X2-X1-0, X2-X1-1, X2-X1-2,
XX X2-X1-3, X2-X1-4, X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and
XX X1-X1-4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
XX X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
XX L, E, S, T, K, H, or E; X1-0 = L, I, V, A, F, M, or K; X1-1 = A, I, V,
XX T, V, N, Q or G; X1-4 = A, I, V, L, F, T, R, E, or G; L1 = linker
XX comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
XX activate the c-Mpl receptor which mediates the activity of endogenous
XX thrombopoietin. The TMPs are useful for increasing the production of
XX platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
XX is useful for treatment of diseases which involve thrombocytopenia, e.g.
XX aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
XX virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 94.8%; Score 182; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 2e-15; Mismatches 0; Gaps 0;
Matches 34; Conservative 0; Indels 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGIEGPTLRQWLAAARA 36
|||||
Db 1 IEGPTLRQWLAAARAGGGGGIEGPTLRQWLAAARA 36

RESULT 7
AAY96525
ID AAY96525 standard; peptide; 36 AA.

XX AAY96525;
XX 04-SEP-2000 (first entry)
XX Thrombopoietin mimetic peptide compound 6.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX Synthetic.

Key	Location/Qualifiers
FT Modified-site 1	/note= "optionally linked to an Fc molecule"
FT Peptide 1..14	/label= TMP_1
FT Peptide 15..18	/label= linker
FT Peptide 19..32	/label= TMP_2
FT Modified-site 32	/note= "optionally linked to an Fc molecule"

XX WO200024770-A2.
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-US24834.
XX 23-OCT-1998; 98US-0105348.
XX

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -
XX
XX Example 1; Page 324; 608pp; English.
XX
XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)-a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1-, (L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions can
CC be used for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAG69443
CC to AAG69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 36 AA;
XX
XX Query Match 100.0%; Score 192; DB 21; Length 36;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-16;
XX Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 IEGPTLQWLAAAGGGGSGGIEGPTLQWLAAARA 36
XX |
XX DB 1 IEGPTLQWLAAAGGGGSGGIEGPTLQWLAAARA 36
XX |
XX RESULT 2
XX AAY95526
XX ID AAY95526 standard; peptide; 36 AA.
XX AC AAY95526;
XX
XX 04-SEP-2000 (first entry)
XX
XX Thrombopoietin mimetic peptide compound 7.
XX
XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
XX anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
XX immunosuppressive; anti-inflammatory; linker.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1
XX Peptide /note= "optionally linked to an Fc molecule"
XX 1..14 /label= TMP_1
XX Peptide 15..18 /label= linker
XX 19..32 /label= TMP_2
XX
XX WO200024770-A2.
XX
XX 04-MAY-2000.
XX
XX 22-OCT-1999; 99WO-US24834.
XX
XX 23-OCT-1998; 98US-0105348.
XX
XX (AMGE-) AMGEN INC.
XX

PI Liu C, Feige U, Cheetham J;
XX WPI; 2000-365108/31.
XX
XX Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia
XX
XX Claim 16; Page 62; 91pp; English.
XX
XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2],
CC is new; TMP-1 and TMP-2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X2-X1-0, X2-X1-1, X2-X1-2,
CC X2-X1-3, X2-X1-4, X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and
CC X1-X1-4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
CC X4 = P; X5 = T or F; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
CC or E; X9 = W, Y or F; X10 = L, I, V, A, F, M, or K; X11 = A, I, V,
CC L, F, S, T, K, H, or E; X12 = A, I, V, L, F, G, S, or O; X13 = R, K,
CC T, V, N, Q or G; X14 = A, I, V, L, F, T, R, E, or G; L1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
XX Sequence 36 AA;
XX
XX Query Match 100.0%; Score 192; DB 21; Length 36;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-16;
XX Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 IEGPTLQWLAAAGGGGSGGIEGPTLQWLAAARA 36
XX |
XX DB 1 IEGPTLQWLAAAGGGGSGGIEGPTLQWLAAARA 36
XX |
XX RESULT 3
XX AAB16963
XX ID AAB16963 standard; Protein; 36 AA.
XX AC AAB16963;
XX
XX 31-OCT-2000 (first entry)
XX
XX TPO-mimetic peptide TMP-TMP SEQ ID NO:14.
XX
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.
XX
XX Synthetic.
XX
XX WO200024782-A2.
XX
XX 04-MAY-2000.
XX
XX 25-OCT-1999; 99WO-US25044.
XX
XX 23-OCT-1998; 98US-0105371.
XX
XX 22-OCT-1999; 99US-0428082.
XX
XX (AMGE-) AMGEN INC.
XX
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI; 2000-350702/30.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 16.1874 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-32

Perfect score: 192

Sequence: 1 IEPTLRQWLAAAGGSGGIEGTPLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	36	21	AA197306
2	192	100.0	36	21	AA196526
3	182	94.8	36	21	AA16963
4	182	94.8	36	21	AA17293
5	182	94.8	36	21	AA17301
6	182	94.8	36	21	AA196523
7	182	94.8	36	21	AA196525
8	182	94.8	41	21	AA196528
9	182	94.8	42	21	AA17281
10	182	94.8	42	21	AA17282
11	182	94.8	42	21	AA17308

12	182	94.8	42	21	AA196530
13	182	94.8	60	21	AA17311
14	182	94.8	269	21	AA16960
15	182	94.8	289	21	AA196531
16	179	93.2	36	21	AA17303
17	179	93.2	36	21	AA17307
18	179	93.2	36	21	AA196524
19	178	92.7	268	21	AA16959
20	171.5	89.3	35	21	AA17292
21	171.5	89.3	37	21	AA17294
22	171	89.1	38	21	AA17295
23	170.5	88.8	39	21	AA17304
24	170.5	88.8	39	21	AA17305
25	170	88.5	40	21	AA17302
26	169	88.0	42	21	AA17296
27	165	85.9	34	21	AA17291
28	164.5	85.7	33	21	AA17290
29	158	82.3	32	21	AA17289
30	156	81.2	36	21	AA17298
31	156	81.2	36	21	AA17297
32	154	80.2	36	21	AA196521
33	154	80.2	36	21	AA196522
34	151.5	78.9	31	21	AA17288
35	145	75.5	30	21	AA17287
36	145	75.5	32	21	AA17297
37	145	75.5	32	21	AA196520
38	145	75.5	34	21	AA196527
39	145	75.5	29	21	AA17286
40	138.5	72.1	29	21	AA17285
41	132	68.8	28	21	AA16970
42	131.5	68.5	29	21	AA16973
43	130.5	68.0	31	21	AA16974
44	130.5	68.0	31	21	AA16971
45	125.5	65.4	29	21	AA16971

ALIGNMENTS

RESULT 1

AA17306

ID AA17306 standard; Peptide; 36 AA.

XX AA17306;

XX 31-OCT-2000 (first entry)

DT TPO-mimetic peptide sequence SEQ ID NO:362.

DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmunity disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

PA Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 29.48; Score 58; DB 10; Length 491;

Wed Oct 9 10:30:11 2002

```

RA RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL
DR EMBL; AE003474; AA547627.1.
DR Flybase; FBgn0035323; CG13807.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHLL.
SQ
SEQUENCE 170 AA; 19099 MW; 477D79D55ADF4CE5 CRC64;
Query Match 29.4%; Score 58; DB 5; Length 170;
Best Local Similarity 45.8%; Pred. No. 22;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 2 EGPTLRQWLAAARAGGCGGGGEG 25
DB 47 EPPIVENWM---GGGGGGGFGQ 66

RESULT 29
Q99034 PRELIMINARY; PRT; 302 AA.
AC Q99034;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ACETYLXYLAN ESTERASE PRECURSOR (EC 3.1.1.72).
GN AXEL.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreales; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 158-186, AND CHARACTERIZATION.
RC STRAIN-RUTC-30;
RX MEDLINE-96235218; PubMed-8647098;
RA Margolles-Clark E., Tenkanen M., Soederlund H., Penttilae M.;
RT "Acetyl xylan esterase from Trichoderma reesei contains an active-site
RT serine residue and a cellulose-binding domain."
RT Eur. J. Biochem. 237:553-560(1996).
RN [2]
RP FUNCTION.
RC STRAIN-RUTC-30;
RA Poutanen K., Sundberg M., Korte H., Puls J.;
RT "Deacetylation of xylans by acetyl esterases of Trichoderma reesei."
RL Appl. Microbiol. Biotechnol. 33:506-510(1990).
RN [3]
RP CHARACTERIZATION.
RC STRAIN-RUTC-30;
RA Sundberg M., Poutanen K.;
RT "Purification and properties of two acetyl xylan esterases of
RT Trichoderma reesei."
RL Biotechnol. Appl. Biochem. 13:11-11(1991).
RN [4]

```

```

RA RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL
DR EMBL; AE003474; AA547627.1.
DR Flybase; FBgn0035323; CG13807.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHLL.
SQ
SEQUENCE 170 AA; 19099 MW; 477D79D55ADF4CE5 CRC64;
Query Match 29.4%; Score 58; DB 5; Length 170;
Best Local Similarity 45.8%; Pred. No. 22;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 2 EGPTLRQWLAAARAGGCGGGGEG 25
DB 47 EPPIVENWM---GGGGGGGFGQ 66

RESULT 29
Q99034 PRELIMINARY; PRT; 302 AA.
AC Q99034;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ACETYLXYLAN ESTERASE PRECURSOR (EC 3.1.1.72).
GN AXEL.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreales; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 158-186, AND CHARACTERIZATION.
RC STRAIN-RUTC-30;
RX MEDLINE-96235218; PubMed-8647098;
RA Margolles-Clark E., Tenkanen M., Soederlund H., Penttilae M.;
RT "Acetyl xylan esterase from Trichoderma reesei contains an active-site
RT serine residue and a cellulose-binding domain."
RT Eur. J. Biochem. 237:553-560(1996).
RN [2]
RP FUNCTION.
RC STRAIN-RUTC-30;
RA Poutanen K., Sundberg M., Korte H., Puls J.;
RT "Deacetylation of xylans by acetyl esterases of Trichoderma reesei."
RL Appl. Microbiol. Biotechnol. 33:506-510(1990).
RN [3]
RP CHARACTERIZATION.
RC STRAIN-RUTC-30;
RA Sundberg M., Poutanen K.;
RT "Purification and properties of two acetyl xylan esterases of
RT Trichoderma reesei."
RL Biotechnol. Appl. Biochem. 13:11-11(1991).
RN [4]

```

```

RA RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL
DR EMBL; AE003474; AA547627.1.
DR Flybase; FBgn0035323; CG13807.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHLL.
SQ
SEQUENCE 170 AA; 19099 MW; 477D79D55ADF4CE5 CRC64;
Query Match 29.4%; Score 58; DB 5; Length 170;
Best Local Similarity 45.8%; Pred. No. 22;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 2 EGPTLRQWLAAARAGGCGGGGEG 25
DB 47 EPPIVENWM---GGGGGGGFGQ 66

RESULT 29
Q99034 PRELIMINARY; PRT; 302 AA.
AC Q99034;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ACETYLXYLAN ESTERASE PRECURSOR (EC 3.1.1.72).
GN AXEL.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreales; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 158-186, AND CHARACTERIZATION.
RC STRAIN-RUTC-30;
RX MEDLINE-96235218; PubMed-8647098;
RA Margolles-Clark E., Tenkanen M., Soederlund H., Penttilae M.;
RT "Acetyl xylan esterase from Trichoderma reesei contains an active-site
RT serine residue and a cellulose-binding domain."
RT Eur. J. Biochem. 237:553-560(1996).
RN [2]
RP FUNCTION.
RC STRAIN-RUTC-30;
RA Poutanen K., Sundberg M., Korte H., Puls J.;
RT "Deacetylation of xylans by acetyl esterases of Trichoderma reesei."
RL Appl. Microbiol. Biotechnol. 33:506-510(1990).
RN [3]
RP CHARACTERIZATION.
RC STRAIN-RUTC-30;
RA Sundberg M., Poutanen K.;
RT "Purification and properties of two acetyl xylan esterases of
RT Trichoderma reesei."
RL Biotechnol. Appl. Biochem. 13:11-11(1991).
RN [4]

```

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL451109; CAC18624.2; -
 KW Hypothetical protein.
 SQ SEQUENCE 776 AA; 82771 MW; C9BEA870D9A437DE CRC64;

Query Match 29.9%; Score 59; DB 3; Length 776;
 Best Local Similarity 53.8%; Pred. No. 75;
 Matches 14; Conservative 3; Mismatches 5; Indels 4; Gaps 2;

QY 15 GGGGGGGI---EG-PTLRQWLAARA 36
 ||| ||||: :| | ||||:|
 Db 678 GGGGGGGVDDGDGPDGFAGWLAQA 703

RESULT 23

ID Q943U7 PRELIMINARY; PRT; 158 AA.
 AC Q943U7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE B1146804.14 PROTEIN.
 GN B1146804.14.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
 RT clone:B1146804.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003205; BAB64583.1; -
 SQ SEQUENCE 158 AA; 14648 MW; 438B72F4B6C86A63 CRC64;

Query Match 29.7%; Score 58.5; DB 10; Length 158;
 Best Local Similarity 61.1%; Pred. No. 18;
 Matches 11; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 9 WLAARAGG-CGGGGIEG 25
 | ||| || ||||: |
 Db 78 WTTARAAGCGCGGAVVG 95

RESULT 24

ID O95692 PRELIMINARY; PRT; 805 AA.
 AC O95692;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE DJ524E15.1 (PEREGRIN (BR140 PROTEIN)) (FRAGMENT).
 GN DJ524E15.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Palmer S.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z84485; CAB06488.1; -
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR001965; PHD.
 DR InterPro; IPR000313; PWMP.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF00628; PHD; 1.
 DR Pfam; PF00855; PWMP; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.

DR SMART; SM00249; PHD; 2.
 DR SMART; SM00293; PWMP; 1.
 DR PROSITE; PSS0014; BROMODOMAIN_2; 2.
 FT NON_TER 1
 FT NON_TER 805
 SQ SEQUENCE 805 AA; 90851 MW; E28C017F5C545334 CRC64;

Query Match 29.7%; Score 58.5; DB 4; Length 805;
 Best Local Similarity 48.1%; Pred. No. 88;
 Matches 13; Conservative 3; Mismatches 8; Indels 3; Gaps 2;

QY 8 QW-LAARAGG-CGGGGIEGPTLRQW 31
 || | || | || | || | || | || |
 Db 695 QWGAASRAFGGSCAGLAGGARRRW 721

RESULT 25

ID Q9JMH4 PRELIMINARY; PRT; 1431 AA.
 AC Q9JMH4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE COLLAGEN TYPE XVII.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;
 RT "Mesocricetus auratus mRNA for type XVII collagen.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB027759; BAA94381.1; -
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 5
 SQ SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9A5C CRC64;

Query Match 29.7%; Score 58.5; DB 11; Length 1431;
 Best Local Similarity 60.0%; Pred. No. 1.6e+02;
 Matches 15; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 12 ARAGGGGGGGIEGPTLRQWLAARA 36
 || ||| |||| | | | | | | | | | |
 Db 438 ARGGGGGGGGGGGT--WGAAPA 459

RESULT 26

ID Q9FU26 PRELIMINARY; PRT; 117 AA.
 AC Q9FU26;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE P0671B11.11 PROTEIN.
 GN P0671B11.11.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0671B11.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002746; BAB12695.1; -
 SQ SEQUENCE 117 AA; 12397 MW; A04617B3DEF9F4B3 CRC64;

Query Match 29.4%; Score 58; DB 10; Length 117;

```
Qy 4 PTLRQW-----LAARAGGCGGGGIEGP 26
   ||| : |||| ||| : ||
Db 219 PRLRGWGSMSRQVGGGRAGSGGGVGLRGP 248

RESULT 19
Q9LWC8 PRELIMINARY; PRT; 125 AA.
AC Q9LWC8:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta.
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
   clone: P0483F08.";
RL EMBL; AP002094; BAA96216.1; -.
KW Hypothetical protein.
SQ SEQUENCE 125 AA; 13396 MW; C609D8D0B07BC505 CRC64;

Query Match 29.9%; Score 59; DB 10; Length 125;
Best Local Similarity 40.5%; Pred. No. 12;
Matches 17; Conservative 2; Mismatches 9; Indels 14; Gaps 2;

Qy 2 EGPTLRQWLAARA-----GGCGGGGIEGPTLRQ 30
   || | ||| : |||| ||| : ||
Db 83 EGAAAR-WRAAKSPARGGRRRGCGGGGRPRRR 123

RESULT 20
Q35392 PRELIMINARY; PRT; 492 AA.
AC Q35392:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FORKHEAD 2.
GN FOXD2 OR MF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MESENCHYME;
RX MEDLINE=98168839; PubMed=9510020;
RA Wu S.-Y., Grindley J., Winnier G.E., Hargett L., Hogan B.L.M.;
RT "Mouse Mesenchyme forkhead 2 (Mf2): expression, DNA binding and
   induction by sonic hedgehog during somitogenesis.";
RL Mech. Dev. 70:3-13(1998).
DR EMBL; AF023915; AAB81275.1; -.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T02492; -.
DR MGD; MGI:1347471; Foxd2.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head.1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 492 AA; 48936 MW; 7F82440F4C435702 CRC64;

Query Match 29.9%; Score 59; DB 11; Length 492;

Best Local Similarity 70.6%; Pred. No. 48;
Matches 12; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 LROWLAARAGGCGGGG 22
   ||| : |||| ||| : ||
Db 384 LRQKLTKDAGGAGGGG 400

RESULT 21
O83436 PRELIMINARY; PRT; 683 AA.
ID O83436:
AC O83436:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0421.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
   Spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65409.1; -.
DR TIGR; TP0421; -.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01436; NHL; 4.
DR Pfam; PF00515; TPR; 1.
KW Complete proteome.
SQ SEQUENCE 683 AA; 74518 MW; F91407FA7094AAD1 CRC64;

Query Match 29.9%; Score 59; DB 16; Length 683;
Best Local Similarity 43.8%; Pred. No. 66;
Matches 14; Conservative 2; Mismatches 12; Indels 4; Gaps 1;

Qy 4 PTLRQWLAARAGGCGGGGIEGPTLRQWLAAR 35
   | : ||| | |||| | |||
Db 74 PLILEWL----GNAYRSGIEGALHGWGAAR 101

RESULT 22
Q9HEA4 PRELIMINARY; PRT; 776 AA.
ID Q9HEA4:
AC Q9HEA4:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN B1A5.200.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
```


DR	InterPro; IPR000209; Peptidase_S8.
DR	InterPro; IPR002884; P_domain.
DR	Pfam; PF01483; P; 1.
DR	Pfam; PF00082; Peptidase_S8; 1.
DR	PRINTS; PR00723; SUBTILISIN.
DR	ProDom; PD000717; P_domain; 1.
DR	PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.
DR	PROSITE; PS00138; SUBTILASE_SER; 1.
KW	Protease; Neuropeptide.
SQ	SEQUENCE 654 AA; 71733 MW; D021D4882293C996 CRC64;
Query Match 30.58; Score 60; DB 5; Length 654;	
Best Local Similarity 48.08; Pred. No. 49;	
Matches	12; Conservative 3; Mismatches 8; Indels 2; Gaps
QY	6 LRQWLAAARGGCG -GGIGIEPTL 28
DB	15 LLHWASAGAGGGAGSGAGLSGPV 39
RESULT 14	
Q14060	PRELIMINARY; PRT; 454 AA.
ID	Q14060
AC	AQ14060;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	COPROPORPHYRINOGEN OXIDASE..
GN	CPX.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID	-9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RA	MEDLINE=95078835; PubMed=7987309;
RX	deFau-Larue M., Martasek P., Grandchamp B.;
RT	"Coproporphyrinogen oxidase: gene organization and description of a
RL	mutation leading to exon 6 skipping.";
RM	Hum. Mol. Genet. 3:1325-1330(1994).
RN	[2]
RP	SEQUENCE OF 101-454 FROM N.A.
RC	TISSUE=PLACENTA;
RA	MEDLINE=94114558; PubMed=8286403;
TX	Taketani S., Kohno H., Furukawa T., Yoshinaga T., Tokunaga R.;
NT	"Molecular cloning, sequencing and expression of cDNA encoding human
RL	coproporphyrinogen oxidase.";
RM	Biochim. Biophys. Acta 1183:547-549(1994).
ENBL	EMBL; Z34531; CAA84292.1; JOINED.
ENBL	Z34803; CAA84292.1; JOINED.
ENBL	Z34804; CAA84292.1; JOINED.
ENBL	Z34805; CAA84292.1; JOINED.
ENBL	Z34806; CAA84292.1; JOINED.
ENBL	Z34807; CAA84292.1; JOINED.
ENBL	Z34808; CAA84292.1; JOINED.
ENBL	D16611; BAA04033.1; JOINED.
InterPro	IPR001260; Coprogen_oxidas.
Pfam	PF01218; Coprogen_oxidas; 1.
PRINTS	PR00073; COPRGNOXDASE.
PROSITE	PS01021; COPROGEN_OXIDASE; 1.
SQ	SEQUENCE 454 AA; 50152 MW; 6EC3D15FD8FD86B5 CRC64;

Wed Oct 9, 10:30:11 2002

us-09-422-838c-31.rspt

```

RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone:P0039A07.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003235; BAB64100.1; -.
SQ SEQUENCE 253 AA; 25568 MW; A963166CE5F97B2B CRC64;

Query Match 31.5%; Score 62; DB 10; Length 253;
Best Local Similarity 51.9%; Pred. No. 11;
Matches 14; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGCGGGGIEGTPLR 29
   ||| : ||| | ||| | | :
Db 80 GPTVGVYVAYRAGAGGGGPRGFALK 106

RESULT 6
ID Q9FT25 PRELIMINARY; PRT; 202 AA.
AC Q9FT25;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0436E04.1 PROTEIN.
DE P0436E04.1.
GN Oryza sativa (Rice).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone:P0436E04.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002818; BAB16319.1; -.
SQ SEQUENCE 202 AA; 19763 MW; BFC2520037F8E274 CRC64;

Query Match 31.2%; Score 61.5; DB 10; Length 202;
Best Local Similarity 38.1%; Pred. No. 10;
Matches 16; Conservative 5; Mismatches 12; Indels 9; Gaps 2;

QY 1 IEPTLRQWLAARAGCGGG-----GGIEPTLRQWLA 34
   : | : | | | | | | | | : | : | : |
Db 94 VVPSRCROTAGR-HGGCGGGRWMAAGGRDGGGCRWWAA 134

RESULT 7
ID Q9SDK6 PRELIMINARY; PRT; 439 AA.
AC Q9SDK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL clone:P0705D01.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF000492; BAA84610.1; -.
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 47297 MW; 533EC240CEA1BA2 CRC64;

Query Match 31.0%; Score 61; DB 10; Length 439;
Best Local Similarity 32.0%; Pred. No. 25;
Matches 16; Conservative 2; Mismatches 18; Indels 14; Gaps 1;

QY 1 IEPTLRQWLAARAGCGGG-----IEPTLRQWLAARA 36
   : | | | | | | | | | | : | | | |
Db 39 LHAPLLRLWPLGGGGGGGGGCGGCGVAVGAVGEARSQRAEA 88

RESULT 8
ID Q9UEA1 PRELIMINARY; PRT; 464 AA.
AC Q9UEA1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA-BINDING PROTEIN NOVA-2 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regalia W., Terry A., Garnes J.,
RA Danganan L., Erler A., Christensen M., Amico-Keller G., Avila J., Liu S.,
RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Brower A., Sanders C., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
RT D19S412.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006540; AAD13116.1; -.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH_TYPE_1.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00013; KH-domain; 3.
DR PRINTS; PR00456; RIBOSOMALP2.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS50084; KH_TYPE_1; 3.
NON_TER 1
FT SEQUENCE 464 AA; 45901 MW; 0B16BAE99C271CC3 CRC64;

Query Match 30.5%; Score 60; DB 4; Length 464;
Best Local Similarity 53.6%; Pred. No. 35;
Matches 15; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 9 WLAARAGCGGGGIEGTPLRQWLAARA 36
   : | | | | | | | | | | : | | |
Db 337 YLGAGAGGAGGGG--GPLVAAAAAAGA 362

RESULT 9
ID Q9UNW9 PRELIMINARY; PRT; 492 AA.
AC Q9UNW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA-BINDING PROTEIN NOVA-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99007301; PubMed=9789075;
RA Yang Y.Y., Yin G.L., Darnell R.B.;
RT "The neuronal RNA-binding protein Nova-2 is implicated as the
RT autoantigen targeted in POMA patients with dementia.";

```



```

DR PRINTS; PR00931; MICOLLPTRASE.
DR SMART; SM00089; PKD; 1.
DR PROSITE; PS00093; PKD; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 865 AA; 92392 MW; 2145740361275F8F CRC64;

Query Match 34.0%; Score 67; DB 2; Length 865;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 WLAARAGCGCGGGGIEGP 26
DB 651 WLAACAAGCGGGTNP 668

RESULT 2
Q90YB6 PRELIMINARY; PRT; 1070 AA.
AC Q90YB6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FORMIN BINDING PROTEIN 11-RELATED PROTEIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Kuribayashi T., Ohashi K.;
RT "Chicken Formin binding protein 11-related protein interacted with
RT Chicken Diaphanous."; to the EMBL/GenBank/DBJ databases.
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057590; BAB68206.1; -.
FT NON_TER 1
SQ SEQUENCE 1070 AA; 119556 MW; A4FA68B7D946BEDF CRC64;

Query Match 33.5%; Score 66; DB 13; Length 1070;
Best Local Similarity 48.1%; Pred. No. 17;
Matches 13; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGCGGGGIEGPTLR 29
DB 148 GPSSRQPLRGLLRGCGGSLSPMR 174

RESULT 3
Q9LGC9 PRELIMINARY; PRT; 360 AA.
AC Q9LGC9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 18, Last annotation update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0462H08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002525; BAB07996.1; -.
DR InterPro; IPR000571; Zf-CCCH.
DR Pfam; PF00642; zf-CCCH; 4.
DR SMART; SM00356; Znf_C3H1; 4.

SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 33.0%; Score 65; DB 10; Length 360;
Best Local Similarity 52.0%; Pred. No. 7.4;
Matches 13; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARAGCGGGGIEG 25
DB 26 LEGPMWRMGLGGGGGGGGGGDG 50

RESULT 4
Q9PVG9 PRELIMINARY; PRT; 431 AA.
AC Q9PVG9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POU-BOX PROTEIN BRAIN-2.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Xue J.X., Zhang W., Fu D.C., He R.Q., Xue Z.G.;
RT "gBrain-2, a POU-box gene expressed in quail embryos.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: WITH OTHER HOMEOBOX PROTEINS.
DR EMBL; AF091043; AAF00040.1; -.
DR HSSP; P14859; IOCT.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000327; POU.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF00157; pou; 1.
DR PRINTS; PR00028; POU00028.
DR PRODOM; PD000583; POU; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00352; POU; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS00071; HOMEOBOX_2; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 431 AA; 43722 MW; 1DC47E53F9ACCTD5 CRC64;

Query Match 32.7%; Score 64.5; DB 13; Length 431;
Best Local Similarity 40.5%; Pred. No. 10;
Matches 17; Conservative 2; Mismatches 6; Indels 17; Gaps 2;

QY 8 QWLAARA-----GGCGGGGIEGPTLRQWLAARA 36
DB 58 QWIAALSHGGPGGGGGGGGGGGGEGAP---WAAAAA 95

RESULT 5
Q943K0 PRELIMINARY; PRT; 253 AA.
AC Q943K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P0039A07.6 PROTEIN.
GN P0039A07.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-31
Perfect score: 197
Sequence: 1 IEPTLRQLAARAGCGGGIEGPTLRQLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_podent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	34.0	865	2	054108 streptomyc
2	66	33.5	1070	13	Q90YB6 gallus gall
3	65	33.0	360	10	Q91GCG oryza sativ
4	64.5	32.7	431	13	Q9PVG9 coturnix co
5	62	31.5	253	10	Q943K0 oryza sativ
6	61.5	31.2	202	10	Q9FTZ5 oryza sativ
7	61	31.0	439	10	Q9SDK6 oryza sativ
8	60	30.5	464	4	Q9UEA1 oryza sativ
9	60	30.5	492	4	Q9UNW9 homo sapien
10	60	30.5	498	4	Q43267 oryza sativ
11	60	30.5	500	5	Q19476 caenorhabdi
12	60	30.5	654	5	Q9VBC7 drosophila
13	60	30.5	654	5	Q9UAE7 drosophila
14	59.5	30.2	454	4	Q14060 homo sapien
15	59.5	30.2	454	4	Q96AF3 homo sapien
16	59.5	30.2	488	16	Q9CCCO mycobacteri

17	59.5	30.2	496	2	Q9AD76 streptomyc
18	59.5	30.2	518	2	Q49843 mycobacteri
19	59	29.9	125	10	Q9LWC8 oryza sativ
20	59	29.9	492	11	Q35392 mus musculu
21	59	29.9	683	16	O83436 treponema p
22	59	29.9	776	3	Q9HEA4 neurospora
23	58.5	29.7	158	10	Q94307 oryza sativ
24	58.5	29.7	805	4	O95692 homo sapien
25	58.5	29.7	1431	11	O9JMH4 mesocricetu
26	58	29.4	117	10	O9FU26 oryza sativ
27	58	29.4	134	2	O56434 thermus the
28	58	29.4	170	5	Q9W033 drosophila
29	58	29.4	302	3	Q99034 trichoderma
30	58	29.4	491	10	Q94LP1 oryza sativ
31	58	29.4	516	10	Q9XEJ0 zea mays (m
32	57.5	29.2	244	11	Q9D384 mus musculu
33	57.5	29.2	302	2	O9S596 myxococcus
34	57.5	29.2	495	16	O33230 mycobacteri
35	57	28.9	76	10	Q9C7W8 arabidopsis
36	57	28.9	377	13	Q9YHD0 petromyzon
37	57	28.9	414	3	Q9HFM0 metarhizium
38	57	28.9	524	4	Q9BZE0 homo sapien
39	57	28.9	529	10	O9ASE5 oryza sativ
40	57	28.9	607	2	Q9L8D4 polyangium
41	57	28.9	612	4	Q9P270 homo sapien
42	57	28.9	651	10	Q9LGW5 oryza sativ
43	57	28.9	1130	4	O75182 homo sapien
44	56.5	28.7	176	17	O9YDB1 aeropyrum p
45	56.5	28.7	207	10	O94IW9 oryza sativ

ALIGNMENTS

RESULT 1

ID	054108	PRELIMINARY;	PRT;	865 AA.
AC	054108;			
DT	01-JUN-1998 (TReMBLrel. 06, Created)			
DT	01-JUN-1998 (TReMBLrel. 06, Last sequence update)			
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)			
DE	SC10A5.17 PROTEIN.			
GN	SC10A5.17.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=1902;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	Murphy L., Harris D.;			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBDJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	Parkhill J., Barrell B.G., Rajandream M.A.;			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBDJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RX	MEDLINE=97000351; PubMed=843436;			
RA	Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,			
RA	Kinashi H., Hopwood D.A.;			
RT	"A set of ordered cosmid and a detailed genetic and physical map for			
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";			
RL	Mol. Microbiol. 21:77-96(1996).			
DR	EMBL; AL021529; CAAL6449.1; -			
DR	MEROPS; M09.001; -			
DR	InterPro; IPR002169; Micollptase.			
DR	InterPro; IPR000601; PKD.domain.			
DR	InterPro; IPR000130; Zn_MTPeptidse.			
DR	Pfam; PF01752; Peptidase_M9; 1.			
DR	Pfam; PF00801; PKD; 1.			

Wed Oct 9 10:30:10 2002

us-09-422-838c-31.rsp

Page 18

Db 96 QWLSPTAAAGGSNGG 112

Search completed: October 9, 2002, 09:00:19
Job time : 5.3831 secs

.....

```
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 33 70 GLY-RICH.
FT DNA_BIND 207 266 HOMEBOX.
FT DOMAIN 264 267 POLY-ALA.
FT VARSPPLIC 271 286 LQHAIGPSGMRSLAE -> SVACTAARPTQA (IN
ISOFORM SIX3A).
FT VARSPPLIC 287 333 MISSING (IN ISOFORM SIX3A).
FT CONFLICT 44 44 G -> GG (IN REF. 1).
FT CONFLICT 118 119 VA -> WP (IN REF. 1).
FT CONFLICT 278 333 PSGMRSLAEPGCPHGSAPSTAASTTSVSSILTERADTG
TSILSVTSSDSECDV -> ERDALPGARLPRLPHARLSRVTVH
GGQPDHOCVQPDGARGHRHFDPLGNLQRLGM (IN REF.
1).
SQ SEQUENCE 333 AA; 35592 MW; 1AD7D3C4388043B9 CRC64;

Query Match 26.4%; Score 52; DB 1; Length 333;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 11 AARAGGCGGGGIEGP 26
I ||||| I
DB 59 AGGAGGGGGGGRAP 74

RESULT 29
FXD3_CHICK STANDARD; PRT; 394 AA.
ID FXD3_CHICK
AC P79772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor genesis) (Winged
helix protein CWH-3).
GN FOXD3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=97141794; PubMed=8988052;
RA Freyaldenhoven B.S., Freyaldenhoven M.P., Tacovoni J.S., Vogt P.K.;
RT "Aberrant cell growth induced by avian winged helix proteins.";
RL Cancer Res. 57:123-129(1997).
CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: U37274; AAC60066.1; -.
DR HSSP: Q63245; 2HEP.
DR TRANSFAC: T02495;
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.

DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 33 70 GLY-RICH.
FT DNA_BIND 207 266 HOMEBOX.
FT DOMAIN 264 267 POLY-ALA.
FT VARSPPLIC 271 286 LQHAIGPSGMRSLAE -> SVACTAARPTQA (IN
ISOFORM SIX3A).
FT VARSPPLIC 287 333 MISSING (IN ISOFORM SIX3A).
FT CONFLICT 44 44 G -> GG (IN REF. 1).
FT CONFLICT 118 119 VA -> WP (IN REF. 1).
FT CONFLICT 278 333 PSGMRSLAEPGCPHGSAPSTAASTTSVSSILTERADTG
TSILSVTSSDSECDV -> ERDALPGARLPRLPHARLSRVTVH
GGQPDHOCVQPDGARGHRHFDPLGNLQRLGM (IN REF.
1).
SQ SEQUENCE 333 AA; 35592 MW; 1AD7D3C4388043B9 CRC64;

Query Match 26.4%; Score 52; DB 1; Length 333;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 11 AARAGGCGGGGIEGP 26
I ||||| I
DB 59 AGGAGGGGGGGRAP 74

RESULT 29
FXD3_CHICK STANDARD; PRT; 394 AA.
ID FXD3_CHICK
AC P79772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor genesis) (Winged
helix protein CWH-3).
GN FOXD3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=97141794; PubMed=8988052;
RA Freyaldenhoven B.S., Freyaldenhoven M.P., Tacovoni J.S., Vogt P.K.;
RT "Aberrant cell growth induced by avian winged helix proteins.";
RL Cancer Res. 57:123-129(1997).
CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: U37274; AAC60066.1; -.
DR HSSP: Q63245; 2HEP.
DR TRANSFAC: T02495;
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.

DR PROSITE: PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 67 70 POLY-ALA.
FT DOMAIN 80 91 POLY-GLY.
FT DOMAIN 100 106 POLY-ALA.
FT DNA_BIND 117 211 FORK-HEAD.
SQ SEQUENCE 394 AA; 40995 MW; 324A4B36B9E31899 CRC64;

Query Match 26.4%; Score 52; DB 1; Length 394;
Best Local Similarity 76.9%; Pred. No. 64;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 13 RAGGGGGGGGIEG 25
I ||||| I
DB 82 RGGGGGGGGGIEG 94

RESULT 30
HKUB_LYCES STANDARD; PRT; 426 AA.
ID HKUB_LYCES
AC Q22300;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Homeobox protein knotted-1 like LET12.
GN LET12.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VENT CHERRY;
RX MEDLINE=98145476; PubMed=9484482;
RA Janssen B.J., Williams A., Chen J.J., Mathern J., Hake S., Sinha N.;
RT "Isolation and characterization of two knotted-like homeobox genes
from tomato.";
RL Plant Mol. Biol. 36:417-425(1998).
CC -!- FUNCTION: MAY HAVE A ROLE TO PLAY IN FORMATIVE EVENTS IN OVULE AND
EMBRYO MORPHOGENESIS.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN THE MATURE PLANT.
CC -!- SIMILARITY: BELONGS TO THE TALE/KNOX FAMILY OF HOMEBOX PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: AF000142; AAC49918.1; -.
DR InterPro: IPR001356; Homeobox.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT DOMAIN 15 24 POLY-GLN.
FT DOMAIN 69 76 POLY-ALA.
FT DOMAIN 140 152 POLY-ASN.
FT DOMAIN 283 287 POLY-ASP.
FT DOMAIN 325 348 ELK DOMAIN.
FT DNA_BIND 349 411 HOMEBOX (TALE-TYPE).
SQ SEQUENCE 426 AA; 47581 MW; 5B52B9E0A34A86BC CRC64;

Query Match 26.4%; Score 52; DB 1; Length 426;
Best Local Similarity 64.7%; Pred. No. 68;
Matches 11; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 8 QWLA--ARAGGCGGGG 22
|||: | |||| |||
```


DR HSP; P09651; IHA1.
 DR FlyBase; FBgn0016978; sRNP70K.
 DR InterPro; IPR000504; SRM.
 DR Pfam; PF00076; rtm: 1.
 DR SMART; SM00360; RM: 1.
 DR PROSITE; PS00102; RM: 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KW Nuclear protein; Ribonucleoprotein; RNA-binding; mRNA processing.
 FT DOMAIN 102 180 RNA-BINDING (RRM).
 FT DOMAIN 254 350 ARG/GLU-RICH (MIXED CHARGE).
 FT CONFLICT 278 278 N -> S (IN REF. 1).
 SQ SEQUENCE 448 AA; 52900 MW; 0DDFB3A39CA72AEB CRC64;
 Query Match 26.6%; Score 52.5; DB 1; Length 448;
 Best Local Similarity 52.4%; Pred. No. 63;
 Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 1;
 QY 5 TLQWLAARAGCGCGG---GG 22
 :: || | ||| || ||
 DB 182 TVRGWLPRLGGLGTRRG 202
 RESULT 27
 PAC4_HUMAN STANDARD; PRT: 969 AA.
 AC P29122; Q15099; Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
 AC Q9UEG7; Q9V4G9; Q9V4H0; Q9V4H1;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like proprotein
 DE convertase 4) (SPC4).
 GN PACE4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).
 RC TISSUE=Hepatoma, and Kidney;
 RX MEDLINE=92075167; PubMed=1741956;
 RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
 Barr P.J.;
 RT "Identification of a second human subtilisin-like protease gene in
 RT the tes/fps region of chromosome 15.";
 RL DNA Cell Biol. 10:757-769(1991).
 [2]
 SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).
 RC TISSUE=Placenta;
 RX MEDLINE=94235049; PubMed=8179631;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms.";
 RL Biochem. Biophys. Res. Commun. 200:943-950(1994).
 [3]
 ERRATUM.
 RX MEDLINE=95071480; PubMed=7980617;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms.";
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
 [4]
 SEQUENCE FROM N.A. (ISOFORM PACE4A-II).
 RC TISSUE=Placenta;
 RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
 Matsuda Y.;
 RT "Identification of a novel PACE4 isoform, PACE4E.";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).
 RP

RC TISSUE=Cerebellum;
 RX MEDLINE=97335942; PubMed=9192737;
 RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
 Akamatsu T., Nagamune H., Matsuda Y.;
 RT "A novel human PACE4 isoform, PACE4E is an active processing protease
 RT containing a hydrophobic cluster at the carboxy terminus.";
 RL J. Biochem. 121:941-948(1997).
 [6]
 SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; A-II; CS; D; E-I; E-II).
 RP MEDLINE=98021085; PubMed=9378725;
 RX Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
 Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
 RA "Genomic organization and alternative splicing of human PACE4 (SPC4),
 RA kexin-like processing endoprotease.";
 RT J. Biochem. 122:438-452(1997).
 [7]
 ALTERNATIVE SPLICING (ISOFORM PACE4CS).
 RP MEDLINE=97064242; PubMed=8906861;
 RX Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
 RA "Functional analysis of human PACE4-A and PACE4-C isoforms:
 RA identification of a new PACE4-CS isoform.";
 RL FEBS Lett. 396:31-36(1996).
 [8]
 CHARACTERIZATION.
 RP MEDLINE=99233559; PubMed=10215603;
 RX Sucic J.F., Moehring J.M., Innocencio N.M., Luchini J.W.,
 Moehring T.J.;
 RA "Endoprotease PACE4 is Ca²⁺-dependent and temperature-sensitive and
 RA can partly rescue the phenotype of a furin-deficient cell strain.";
 RT Biochem. J. 339:639-647(1999).
 [9]
 PROCESSING.
 RP MEDLINE=98408849; PubMed=9738469;
 RX Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
 Tsuji A., Matsuda Y.;
 RA "Biosynthetic processing and quaternary interactions of proprotein
 RA convertase SPC4 (PACE4)."
 RT FEBS Lett. 434:155-159(1998).
 [10]
 FUNCTION: LIKELY TO REPRESENT AN ENDOPEPTIDASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS,
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- COFACTOR: PACE4A IS PROBABLY CALCIUM-DEPENDENT.
 CC -1- SUBUNIT: THE PACE4A-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE
 CC RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX
 CC WHEREAS MATURE PACE4A-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT
 CC PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: PACE4A-I AND PACE4A-II ARE SECRETED. PACE4C
 CC AND PACE4CS ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM
 CC IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED
 CC INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-
 CC TERMINUS. PACE4B MIGHT BE SECRETED.
 CC -1- ALTERNATIVE PRODUCTS: 8 ISOFORMS; PACE4A-I/PACE4 (SHOWN HERE),
 CC PACE4A-II, PACE4B/PACE4.1, PACE4C, PACE4CS, PACE4D, PACE4E-I AND
 CC PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B,
 CC C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.
 CC -1- TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE
 CC RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN,
 CC PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT
 CC COMPARATIVELY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST
 CC EXPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC
 CC KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE
 CC EXPRESSED IN PLACENTA. PACE4E-I IS EXPRESSED IN CEREBELLUM,
 CC PLACENTA AND PITUITARY. PACE4E-II IS AT LEAST PRESENT IN
 CC CEREBELLUM.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM. ISOFORM PACE4D LACKS THE PROPEPTIDE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.

```

QY 4 PTLRQWLAARAGGG---CGGGGIEGPTLR 29
  I I I I I I I I I I I I I I I I
Db 27 PIARQRCSAAGGNWYPVGGGIGDPWCR 55

RESULT 25
CYB_MICIK
ID CYB_MICIK STANDARD; PRT; 370 AA.
AC Q9MLK2;
DT 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome B.
GN MTCYB OR COB OR CYTB.
OS Micropechis ikaheka.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Notochidae; Micropechis.
OX NCBI_TaxID=66188;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20229584; PubMed=10764543;
RA Slowinski J.B., Keogh J.S.;
RT "Phylogenetic relationships of elapid snakes based on cytochrome b
  mtDNA sequences."
RL Mol. Phylogenet. Evol. 15:157-164(2000).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF217826; AAF37245.1;
CC InterPro: IPR000179; Cytochrome_b_c1.
CC Pfam: PF00032; cytochrome_b_c1.
CC DR PROSITE: PS00193; CYTOCHROME_B_N; 1.
CC DR PROSITE: PS00192; CYTOCHROME_B_QQ; 1.
CC KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
CC Heme.
CC FT METAL 75 75 IRON 1 (HEME B562 AXIAL LIGAND).
CC FT METAL 89 89 IRON 2 (HEME B566 AXIAL LIGAND).
CC FT METAL 174 174 IRON 2 (HEME B562 AXIAL LIGAND).
CC FT METAL 188 188 IRON 1 (HEME B566 AXIAL LIGAND).
CC SQ SEQUENCE 370 AA; 42083 MW; CCDE45269CAB2B9D CRC64;

Query Match 26.6%; Score 52.5; DB 1; Length 370;
Best Local Similarity 41.9%; Pred. No. 53;
Matches 13; Conservative 2; Mismatches 9; Indels 7; Gaps 1;

QY 3 GPTLRQWLAARAGGGCGGGGIEGPTLRQWLA 33
  I I I I I I I I I I I I I I I I
Db 149 GPTLTWL-----WGGFSINDPLTRFFA 172

RESULT 26
RUI7_DROME
ID RUI7_DROME STANDARD; PRT; 448 AA.
AC P17133; O9VM56;
DT 01-AUG-1990 (Rel. 15, Created)

```

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE U1 small nuclear ribonucleoprotein 70 kDa (UI snRNP 70 kDa) (snRNP70).
GN SNRNP70K OR SNRNP27D OR CG8749.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258833; PubMed=1692955;
RA Mancebo R., Lo P.C.H., Mount S.M.;
RT "Structure and expression of the Drosophila melanogaster gene for the
  U1 small nuclear ribonucleoprotein particle 70K protein."
RL Mol. Cell. Biol. 10:2492-2502(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: MEDIATES THE SPLICING OF PRE-MRNA BY BINDING TO THE STEM
CC LOOP I REGION OF U1-SNRNA.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31162; AAA28859.1;
CC DR EMBL; AE003615; AAF52471.1;
CC PIR: A36311; A36311.

```

SEQUENCE FROM N.A.
STRAIN=SPRAGUE-DAWLEY;
MEDLINE=96305355; PubMed=8688467;
X Flowers K.M., Mellor H., Matts R.L., Kimball S.R., Jefferson L.S.;
T "Cloning and characterization of complementary and genomic DNAs
T encoding the epsilon-subunit of rat translation initiation
T factor-2B";
T Blochim. Biophys. Acta 1307:318-324(1996).
C -1- FUNCTION: CATALYZES THE EXCHANGE OF EUKARYOTIC INITIATION FACTOR
C 2- BOUND GDP FOR GTP.
C -1- SUBUNIT: COMPLEX OF FIVE DIFFERENT SUBUNITS; ALPHA, BETA, GAMMA,
C DELTA AND EPSILON.
C -1- SIMILARITY: BELONGS TO THE EIF-2B GAMMA/EPSILON SUBUNITS FAMILY.

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
C or send an email to license@isb-sib.ch).

C EMBL; U19516; AAB17690.1; -;
C EMBL; U19511; AAB17691.1; -;
C InterPro; IPR001451; Hexapep_transf.
C InterPro; IPR003307; eIF5C.
C Pfam; PF001132; hexapep; 3.
C SMART; SM00515; eIF5C; 1.
C Amino-acid biosynthesis; Translation regulation.
C DOMAIN 19 26 POLY-GLY.
C DOMAIN 34 37 POLY-PRO.
C SEQUENCE 716 AA; 80240 MW; C6E4BFC0E06AF6F1 CRC64;

C Query Match 26.9%; Score 53; DB 1; Length 716;
C Best Local Similarity 43.3%; Pred. No. 83;
C Matches 13; Conservative 3; Mismatches 8; Indels 6; Gaps 1;

QY 11 AARAGGCGGGGIEG-----PTLRQWLAA 34
| : ||| ||||| | | : |
DB 15 ANKRGSGGGGTGAGEEPPPLQAVLA 44
| : ||| ||||| | | : |

RESULT 23
SSB_RHOSH STANDARD; PRT; 174 AA.
Q92A08.
30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
Single-strand binding protein (SSB) (Helix-destabilizing protein).
SSB
Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
NCBI_TaxID=1063;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RA Zellstra-Rvalls J.H., Kaplan S.;
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
CC CHROMOSOME. IT IS ALSO INVOLVED IN DNA RECOMBINATION AND REPAIR
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SSB FAMILY.

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
C or send an email to license@isb-sib.ch).

QY 12 ARAGGCGGGGIEB--GP 26
| |||| ||||| | | |
DB 122 AGAGGGGMMGGGYEDRGCP 139
| |||| ||||| | | |

RESULT 24
SPIN_CBEVP STANDARD; PRT; 341 AA.
ID P23061;
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Spindolin precursor (Spheroidin).
GN P50 OR SPH.
OS Choristoneura biennis entomopoxvirus (CbEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=10288;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 21-50.
RP MEDLINE=90223988; PubMed=2327073;
RX Yuen L., Dionne J., Arif B., Richardson C.;
RA "Identification and sequencing of the spheroidin gene of
RT Choristoneura biennis entomopoxvirus";
RL Virology 175:427-433(1990).
[2]
RP REVISION TO FUNCTION.
RX MEDLINE=93389435; PubMed=8376960;
RA Dall D., Sriskantha A., Vera A., Lal-Fook J., Symonds T.;
RA "A gene encoding a highly expressed spindle body protein of Heliothis
RT armigera entomopoxvirus";
RL J. Gen. Virol. 74:1811-1818(1993).
CC -1- FUNCTION: THIS PROTEIN IS A SPINDLE BODY PROTEIN.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: WITH HAEPV SPINDOLIN AND ACMNPV SPINDOLIN-LIKE
CC PROTEIN.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A SPHEROIDIN.

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
C or send an email to license@isb-sib.ch).

C EMBL; M34140; AAA42887.1; -;
C InterPro; IPR004302; Chitin_bind_3.
C Pfam; PF03067; Chitin_bind_3; 1.
C Signal; Late protein.
FT SIGNAL 1 20
FT CHAIN 21 341 SPINDOLIN.
SQ SEQUENCE 341 AA; 38709 MW; E84EF9BCD901E72F CRC64;

C Query Match 26.6%; Score 52.5; DB 1; Length 341;
C Best Local Similarity 44.8%; Pred. No. 50;
C Matches 13; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

Matches 17; Conservative 2; Mismatches 12; Indels 17; Gaps 2;

QY 6 LRLWLAARAGG-CGGGGTGGPTLR-----OWLAARA 36
 |||||
 Db 33 LRSWLLSRQPAETGGGQGGPGLRLITGLFGAGLGGAWLALRA 80
 |||||

RESULT 20

SYK_AERPE
 ID SYK_AERPE STANDARD; PRT; 562 AA.
 AC Q9YFT9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine-tRNA ligase) (LYSRS).
 GN LYSS OR AP00161.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 ON NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka H.,
 Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1";
 RL DNA Res. 6:83-101(1999).
 CC -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
 CC + L-lysyl-tRNA(Lys).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP000058; BAA79072.1; -
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002904; tRNA-synt_lys_1.
 DR Pfam; PF01921; tRNA-synt_1f; 1.
 DR PROSITE; PS00178; AA-TRNA_LIGASE_I; FALSE_NEG.
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 50 58 "HIGH" REGION.
 FT SITE 305 309 "KMSKS" REGION.
 SQ SEQUENCE 562 AA; 65114 MW; 753664E2937FBF27 CRC64;
 Query Match 27.28; Score 53.5; DB 1; Length 562;
 Best Local Similarity 39.3%; Pred. No. 60;
 Matches 11; Conservative 4; Mismatches 10; Indels 3; Gaps 1;

QY 8 OWLAARAGG---CGGGGTGGPTLRWL 32
 |||||
 Db 293 EWSVLRAGGREADMSGGTGTITPREWL 320
 |||||

RESULT 21

GATA_MYCLE
 ID GATA_MYCLE STANDARD; PRT; 497 AA.
 AC G3105;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Translation initiation factor eIF-2B epsilon subunit (eIF-2B GDP-GTP
 exchange factor).
 GN EIF2B5 OR EIF2BE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]

DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
 subunit A).
 GN GATA OR ML1702 OR MLCB637.13.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 Barrell B.G.;
 "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
 CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISCHARGED GLU-
 CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
 CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
 CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
 CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
 CC -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z99263; CAB16428.1; -
 DR EMBL; AL583923; CAC30655.1; -
 DR Leproma; ML1702; -
 DR InterPro; IPR0000120; Amidase.
 DR Pfam; PF01425; Amidase; 1.
 DR PROSITE; PS00571; AMIDASES; 1.
 KW Protein biosynthesis; Ligase; Complete proteome.
 SQ SEQUENCE 497 AA; 51536 MW; D3723D871518BDC7 CRC64;

Query Match 26.9%; Score 53; DB 1; Length 497;
 Best Local Similarity 52.6%; Pred. No. 61;
 Matches 10; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 GPTLRWLAAARAGGGGGG 21
 |||||
 Db 145 GPTLRWVWVDRVPGSGGG 163
 |||||

RESULT 22

EZBE_RAT
 ID EZBE_RAT STANDARD; PRT; 716 AA.
 AC Q64350;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Translation initiation factor eIF-2B epsilon subunit (eIF-2B GDP-GTP
 exchange factor).
 GN EIF2B5 OR EIF2BE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]

Burtis K.C., Busam D.A., Butler H.H., Cadieu E.E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W., Flockerzi A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwag C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z., Liang Y., Lin X., Liu X., Lele Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Murphy B., Murphy L., Murzyn D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclab J.M., Palazzolo M., Pittman G.S., Pan S., Pollard R.J., Purf V., Reese M.G., Reinert K., Remington K.I., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R., Ventner E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhao X., Zhang G., Zhao Q., Zheng L., Zhong X.H., Zhong F.N., Zhou X., Zhu S., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
The genome sequence of *Drosophila melanogaster*.;
Science 287:2185-2195(2000).
RT SCIENCE BACKGROUND POTASSIUM CHANNEL. RECTIFICATION IS DEPENDENT ON EXTERNAL POTASSIUM CONCENTRATION. ACTS AS AN OUTWARDLY RECTIFYING CHANNEL BUT AS EXTERNAL POTASSIUM LEVELS INCREASE, THIS IS REVERSED.
CC -! SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -! TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN ADULT, STRONGEST EXPRESSION IN MUSCLE, BRAIN AND OVARY. ALSO PRESENT AT LOW LEVELS IN LARVA AND EMBRYO.
CC -! MISCELLANEOUS: INHIBITED BY BARIUM.
CC -! SIMILARITY: BELONGS TO THE TWO PORE DOMAIN FAMILY OF POTASSIUM CHANNELS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; U55321; AAC69250.1; --
DR EMBL; AE003484; AAFA7972.1; --
DR FlyBase; FBgn0017561; Orkl.
DR InterPro; IPRO03280; 2porek_channel.
DR InterPro; IPRO01622; Channel_pore_k.
DR Pfam; PF02034; TWIK_Channel; 1.
DR PRINTS; PR01333; 2POREKCHANEL.
KW Ionic channel; Transmembrane; Ion transport; Potassium transporter
KW Glycoprotein.
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 95 111 PORE-FORMING 1 (POTENTIAL).
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 141 170 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 171 191 POTENTIAL.
FT TRANSMEM 208 224 PORE-FORMING 2 (POTENTIAL).
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 265 1001 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 58 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD SEQUENCE 1001 AA; 109289 MW; 09AELN3669072E07 CRC64;

Query Match 27.4%; Score 54; DB 1; Length 1001;
Best Local Similarity 52.2%; Pred. No. 87;
Matches 12; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 8 OWLAARAGGGGCGGGIEGTTLRQ 30
DB 761 QQAAAAAGGAAGGGSIRGRKQ 783
PRT: 266 AA.
STANDARD:
SC02_HUMAN
ID SC02_HUMAN
DC 043819; Q9UK87;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SC02 protein homolog, mitochondrial precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUP=Monocytes;
RA Snink L.J., Burton J.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
[2]
RM MEDLINE=20014747; PubMed=10545952;
RX Papadopoulou L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I., RA Sadlock J.E., Krishna S., Walker W., Selby J., Clerum D.M., RA Van Coster R., Lyon G., Scalsis E., Lebel R., Kaplan P., Shanske S., RA De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.; and
RT "Fatal infantile cardioencephalopathy with COX deficiency and mutations in SC02, a COX assembly gene.";
Nat. Genet. 23:333-337(1999).
RL FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
CC -! SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -! TISSUE SPECIFICITY: UBUIQUITOUS.
CC -! DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE CARDIOENCEPHALOPATHY WITH COX DEFICIENCY. THIS DISEASE IS CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND GLISTOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX DEFICIENCIES.
CC -! SIMILARITY: BELONGS TO THE SC01/2 FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AF177385; AAF05313.1; --
DR EMBL; AL021683; CAA16671.1; --
DR MM; 604272; --
DR MM; 604377; --
DR MM; 220110; --
DR InterPro; IPR003782; SC01_SenC.
DR Pfam; PF02630; SC01-SenC; 1.
KW Mitochondrion; Transit peptide; Disease mutation; Polymorphism.
FT TRANSIT 1 41 MITOCHONDRION (POTENTIAL).
FT CHAIN 42 266 SC02 PROTEIN HOMOLOG.
FT VARIANT 20 20 R -> P (IN DBSNP:140523).
FT VARIANT 140 140 FTID-VAR_011738.
FT VARIANT 225 225 E -> K (IN PIC).
FT VARIANT 225 225 /FTID-VAR_008874.
FT VARIANT 225 225 S -> F (IN PIC).
FT VARIANT 225 225 /FTID-VAR_008875.
FT SEQUENCE 266 AA; 29810 MW; BC2F40E057329BF3 CRC64;

Query Match 27.2%; Score 53.5; DB 1; Length 266;
Best Local Similarity 35.4%; Pred. No. 31;

DE Hypothetical 20.2 kDa protein in RNA2.
 OS Tomato ringspot virus (isolate raspberry) (Tomrsv).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
 CC Nepovirus.
 CC NCBI_TaxID=12281;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91311402; PubMed=1856689;
 RA Rott M.E., Tremaine J.H., Rochon D.M.;
 RT "Nucleotide sequence of tomato ringspot virus RNA-2";
 RL J. Gen. Virol. 72:1505-1514 (1991).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D12477; BAA02044.1; -
 DR PIR; JQ1094; JQ1094.
 DR HSSP; P04002; LWFA.
 KW Hypothetical protein.
 FT DOMAIN 15 22 POLY-GLY.
 FT DOMAIN 61 66 POLY-GLY.
 FT DOMAIN 144 148 POLY-GLY.
 FT SEQUENCE 201 AA; 20194 MW; 9038506E18D7B450 CRC64;
 SQ
 Query Match 27.4%; Score 54; DB 1; Length 201;
 Best Local Similarity 57.7%; Pred No. 22;
 Matches 15; Conservative 1; Mismatches 6; Indels 4; Gaps 1;
 QY 13 RAGGCGGGGIE-----GPTLRQWLAA 34
 DB 13 RAGGCGGGGKEVFKAGRTLLKVLKA 38
 RESULT 17
 ID Y870_METJA STANDARD; PRT; 620 AA.
 AC Q58280;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0870.
 GN MJ0870.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 CC NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii";
 RL Science 273:1058-1073 (1996).
 CC -!- SIMILARITY: TO COENZYME F420 HYDROGENASE BETA SUBUNIT.
 CC -!- SIMILARITY: TO M. JANNASCHII MJ1349, MJ0725 AND MJ0551.
 CC -!- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
 CC FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
 CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U67531; AAB98876.1; -
 DR HSSP; Q45560; IBOX.
 DR TIGR; MJ0870; -
 DR InterPro; IPR001450; 4Fe4S_ferredoxin.
 DR Pfam; IPR000660; Nir_Sir.
 DR Pfam; PF00037; fer4; 3.
 DR PRINTS; PR00397; SIROHAEM.
 DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
 DR PROSITE; PS00365; NIR_Sir; 1.
 KW Hypothetical protein; Oxidoreductase; Heme; Iron-sulfur; 4Fe-4S;
 KW Complete proteome.
 FT METAL 428 428 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 434 434 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 468 468 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 472 472 IRON-SULFUR (4FE-4S) AND SIROHEME
 FT (BY SIMILARITY).
 SQ SEQUENCE 620 AA; 69793 MW; 9D71D2580D7D0BA8 CRC64;
 Query Match 27.4%; Score 54; DB 1; Length 620;
 Best Local Similarity 43.5%; Pred. No. 58;
 Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
 QY 2 EGPTLRQWLAA RAGCGGGGIE 24
 DB 418 EGPLVRATLACPGGNCSSGLVD 440
 RESULT 18
 ORKL_DROME
 ID ORKL_DROME STANDARD; PRT; 1001 AA.
 AC Q94526;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Open rectifier potassium channel protein 1 (Two pore domain potassium
 DE channel Orkl).
 DE ORKL OR C61615.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Larva;
 RX MEDLINE=97075152; PubMed=8917578;
 RA Goldstein S.A.N., Price L.A., Rosenthal D.N., Pausch M.H.;
 FT "Orkl, a potassium-selective leak channel with two pore domains
 RT cloned from Drosophila melanogaster by expression in Saccharomyces
 RL cerevisiae";
 RN Proc. Natl. Acad. Sci. U.S.A. 93:13256-13261 (1996).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananadis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaize R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.B.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

Rhabditophora; Eulecithophora; Revertospermatia; Mediofusata;
Nemertea; Trematoda; Digenea; Strigeidae; Schistosomatoidea;
Schistosomatidae; Schistosoma.
NCBI_TaxID=6182;
[1]
SEQUENCE FROM N.A.
MEDLINE=98318804; PubMed=2457805;
Hedstrom R., Culpepper J., Schinski V., Agabian N., Newport G.;
"Schistosome heat-shock proteins are immunologically distinct
host-like antigens."; Mol. Biochem. Parasitol. 29:275-282(1988).
-!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M21011; AAA29897.1; .
PIR; A54507; A54507.
HSP; P08109; ICKR.
InterPro; IPR001023; HSP70.
PROSITE; PS00297; HSP70_1; PARTIAL.
PROSITE; PS00329; HSP70_2; PARTIAL.
PROSITE; PS01036; HSP70_3; PARTIAL.
ATP-binding; Heat shock.
FT NON_TER 1
KW FT SEQUENCE 198 AA; 21845 MW; 800F9586046D5313 CRC64;
SQ
Query Match 27.7%; Score 54.5; DB 1; Length 198;
Best Local Similarity 41.4%; Pred. No. 19;
Matches 12; Conservative 3; Mismatches 3; Indels 11; Gaps 1;
QY 13 RAGG-----CCGGGGIEGPTLRQ 30 | |||| :||| :
DB 168 RAGGVPGMPGCGAGGGGGKGPTEE 196 | |||| :||| :
RESULT 15
OVO_DROME OVO_DROME STANDARD; PRT; 1028 AA.
ID F51521; Q9XZU4; PRT; 1028 AA.
AC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ovo protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pserygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophilla.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP RP
RC TISSUE-Ovary;
RX MEDLINE=95021209; PubMed=7935398;
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
melanogaster: relationship to genetic complexity."; Mol. Cell. Biol. 14:6809-6818(1994).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Marvel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required
for female germ line development."; EMBO J. 10:3259-3266(1991).
RL EMBO J. 10:3259-3266(1991).
CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CELLS. PLAYS A ROLE IN GERMLINE SEX DETERMINATION.

```

Best Local Similarity 32.6%; Pred. No. 50;
Matches 15; Conservative 0; Mismatches 9; Indels 22; Gaps 1;

QY 3 GPTLRQW-----LAARAGCGCGGGGIEGP 26
    | | | | |
Db 607 GKTLESWSLCTRCWASKGAAGVGGAGATAGGGGPGGGGGGP 652
    | | | | |

RESULT 12
SYA_CHLMU
ID SYA_CHLMU STANDARD; PRT; 875 AA.
AC Q9PLH5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ALANYL-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
GN ALAS OR TC0125
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."; 28.1397-1406(2000).
RL Nucleic Acids Res.
CC -1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-tRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002279; AAF39003.1; -
CC TIGR; TC0125; -
CC InterPro; IPR002106; AA_CRNA_ligase_II.
CC InterPro; IPR003156; DHHA1.
CC InterPro; IPR002318; tRNA-synt_2c.
CC Pfam; PF02272; DHHA1; 1.
CC Pfam; PF01411; tRNA-synt_2c; 1.
CC PRINTS; PR00980; TRNASYNTHALA.
CC PROSITE; PS00179; AA_TRNA_LIGASE_II.1; FALSE_NEG.
CC PROSITE; PS00339; AA_TRNA_LIGASE_II.2; 1.
CC Aminoacyl-tRNA synthetase; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 875 AA; 90185 MW; A75F8977A23DC41D CRC64;

Query Match 27.9%; Score 55; DB 1; Length 875;
Best Local Similarity 28.6%; Pred. No. 61;
Matches 14; Conservative 5; Mismatches 16; Indels 14; Gaps 1;

QY 1 IEPTLRQWLAARAGCGGGGIE-----GPTLRQWLAAR 35
    | | | | |
Db 825 IDAQLLELLAPYGRGCKGKAVSAQSSKELPQIEVLNTILROWISTR 873
    | | | | |

RESULT 13
LSR2_MYCTU
ID LSR2_MYCTU STANDARD; PRT; 112 AA.
AC O06285;
DT 30-MAY-2000 (Rel. 39, Created)

```

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE LSR2 protein precursor.
GN LSR2 OR RV3597C OR MT3704 OR MTCY07H7B.25.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Krohn A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Deicher A., Utterback T., Weidman J., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DOMINANT T-CELL ANTIGEN AND STIMULATES
CC LYMPHOPROLIFERATION (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z95557; CAB08947.1; -
CC EMBL; AE007170; AAK48061.1; -
CC TIGR; MT3704; -
CC TuberculList; RV3597C; -
CC Antigen; Immune response; Signal; Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 112 LSR2 PROTEIN.
SQ SEQUENCE 112 AA; 12098 MW; A4B32E478CBAC3E4 CRC64;

Query Match 27.7%; Score 54.5; DB 1; Length 112;
Best Local Similarity 33.3%; Pred. No. 12;
Matches 13; Conservative 6; Mismatches 7; Indels 13; Gaps 2;

QY 6 LRQWLAA-----RAGCGGGGI---EGPTLRQW 31
    | | | | |
Db 48 LKQWVAAGRRVGRRRGRSGRGRGAIDREQSAIREW 86
    | | | | |

RESULT 14
HS70_SCHJA
ID HS70_SCHJA STANDARD; PRT; 198 AA.
AC P12795;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Heat shock 70 kDa protein (HSP70) (Fragment).
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;

```

Forman-Kay J.D.;
 "Structure of a Numb prb domain-peptide complex suggests a basis for
 diverse binding specificity.";
 Nat. Struct. Biol. 5:1075-1083(1998).
 CC -!- FUNCTION: NUMB IS REQUIRED IN DETERMINATION OF CELL FATE DURING
 CC SENSORY ORGAN FORMATION IN DROSOPHILA EMBRYOS. IT FUNCTIONS IN
 CC NUCLEI AND SEEMS TO INTERACT WITH NUCLEIC ACIDS.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M27815; AAA28730.1; -;
 CC PIR; A32466; A32466.
 CC PDB; 2NMB; 04-NOV-98.
 CC FlyBase; FBgn0002973; numb.
 CC InterPro; IPR000050; PID_Domain.
 CC Pfam; PF00640; PID; 1.
 CC SMART; SM00462; PTB; 1.
 CC PROSITE; PS01179; PID; 1.
 CC Nucleic protein; ATP-binding; Alternative initiation; 3D-structure.
 KW CHAIN 1 556 NUMB PROTEIN, ZYGOTIC ISOFORM.
 FT CHAIN 42 556 NUMB PROTEIN, MATERNAL ISOFORM
 FT CHAIN 42 556 (PROBABLE).
 FT INIT_MET 42 42 FOR MATERNAL ISOFORM.
 FT NP_BIND 22 29 ATP (POTENTIAL).
 FT DOMAIN 25 57 ARG/LYS-RICH (BASIC).
 FT DOMAIN 81 208 PID.
 SQ SEQUENCE 556 AA; 60628 MW; 4FECAAE9C98FEE71 CRC64;
 Query Match 27.9%; Score 55; DB 1; Length 556;
 Best Local Similarity 42.3%; Pred. No. 41;
 Matches 11; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 8 QWLAARAGGCGGGTGGTPTLRWL 33
 DB 486 QTLASGTGAAGVGGGPDGPDFAEWA 511
 RESULT 11
 ID FZD8_HUMAN STANDARD; PRT; 694 AA.
 AC Q9H461;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Frizzled 8 precursor (Frizzled-8) (Fz-8) (hfz8).
 GN FZD8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21192958; PubMed=11295046;
 SA Saitoh T., Hirai M., Katoh M.;
 RT "Molecular cloning and characterization of human Frizzled-8 gene on
 RT chromosome 10p11.2";
 RL Int. J. Oncol. 18:991-996(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Heath P.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors
 CC are coupled to the beta-catenin canonical signaling pathway, which
 CC leads to the activation of dishevelled proteins, inhibition of GSK-
 CC 3 kinase, nuclear accumulation of beta-catenin and activation of

Wnt target genes. A second signaling pathway involving PKC and
 calcium fluxes has been seen for some family members, but it is
 not yet clear if it represents a distinct pathway or if it can be
 integrated in the canonical pathway, as PKC seems to be required
 for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem
 to involve interactions with G-proteins. May be involved in
 transduction and intercellular transmission of polarity
 information during tissue morphogenesis and/or in differentiated
 tissues.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Most abundant in fetal kidney, followed by
 CC brain and lung. In adult tissues, expressed in kidney, heart,
 CC pancreas and skeletal muscle.
 CC -!- DOMAIN: Lys-Thr-X-X-Trp motif is involved in the activation of
 CC the Wnt/beta-catenin signaling pathway (By similarity).
 CC -!- DOMAIN: The fz domain is involved in binding with Wnt ligands (by
 CC similarity).
 CC -!- SIMILARITY: BELONGS TO FAMILY FZ/SMO OF G-PROTEIN COUPLED
 CC RECEPTORS.
 CC -!- SIMILARITY: CONTAINS 1 FRIZZLED (FZ) DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB043703; BAB41064.1; -;
 CC EMBL; ALI21749; CAC10185.1; -;
 CC MIM; 606146; -;
 CC InterPro; IPR000539; Frizzled.
 CC InterPro; IPR000024; Fz_domain.
 CC InterPro; IPR000832; GPCR_secretin.
 CC Pfam; PF01534; Frizzled; 1.
 CC Pfam; PF01392; Fz; 1.
 CC PRINTS; PR00489; FRIZZLED.
 CC SMART; SM00063; FRI; 1.
 CC PROSITE; PS50038; FZ; 1.
 CC PROSITE; PS50261; G-PROTEIN_RECEP_F2_4; 1.
 KW Multigene family; G-protein coupled receptor; Transmembrane;
 KW Developmental protein; Glycoprotein; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 694 FRIZZLED 8.
 FT DOMAIN 28 275 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 276 296 1 (POTENTIAL).
 FT DOMAIN 297 312 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 313 333 2 (POTENTIAL).
 FT DOMAIN 334 396 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 397 417 3 (POTENTIAL).
 FT DOMAIN 418 439 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 440 460 4 (POTENTIAL).
 FT DOMAIN 461 483 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 484 504 5 (POTENTIAL).
 FT DOMAIN 505 532 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 533 553 6 (POTENTIAL).
 FT DOMAIN 554 584 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 585 605 7 (POTENTIAL).
 FT DOMAIN 606 694 CYTOPLASMIC (POTENTIAL).
 FT FZ. 30 151
 FT DOMAIN 168 172 POLY-PRO.
 FT DOMAIN 194 202 POLY-GLY.
 FT DOMAIN 211 216 POLY-GLY.
 FT DOMAIN 639 663 POLY-GLY.
 FT SITE 608 613 LYS-THR-X-X-X-TRP MOTIF.
 FT SITE 692 694 PDZ-BINDING.
 FT CARBOHYD 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 694 AA; 73300 MW; E740CBFDA2A233EF CRC64;
 Query Match 27.9%; Score 55; DB 1; Length 694;

FT CHAIN	83	103	ANALYSIS: SUMMER

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF042832; AAC15421.1; -;
DR HSP; Q63245; 2HFH.
DR TRANSFAC; T02485; -;
DR MIM; 602211; -;
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS0039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 90 94 POLY-ALA.
FT DOMAIN 101 104 POLY-ALA.
FT DOMAIN 126 217 FORK-HEAD.
FT DOMAIN 247 250 POLY-ALA.
FT DOMAIN 296 306 POLY-ALA.
FT DOMAIN 398 409 POLY-GLY.
FT DOMAIN 421 426 POLY-GLY.
FT DOMAIN 442 445 POLY-ALA.
SQ SEQUENCE 497 AA; 49007 MW; EAAF498D216BE019 CRC64;

Query Match 30.5%; Score 60; DB 1; Length 497;

Best Local Similarity 66.7%; Pred. No. 11;

Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 4 PT--LRQWLAARAGCGCGGG 22

II IIII IIII IIII

Db 385 PTALLRQGLKTDAGGAGGGG 405

RESULT 4

ID SYA_CHLTR STANDARD; PRT; 875 AA.
AC O84754;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine-tRNA ligase) (AlaRS).
GN ALAS OR CT749.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/OW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

EMBL; AE001346; AAC68344.2; -

InterPro; IPR002106; AA_tRNA_ligase_II.

InterPro; IPR003156; DHAL.

DR InterPro; IPR002318; tRNA-synt_2c.

DR Pfam; PF02272; DHAL1; 1.
DR Pfam; PF01411; tRNA-synt_2c; 1.
DR PRINTS; PR00980; TRNASYNTHALA.
DR PROSITE; PS00179; AA_tRNA_LIGASE_II_1; FALSE_NEG.
DR PROSITE; PS00339; AA_tRNA_LIGASE_II_2; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 875 AA; 97671 MW; 81C2DA7B29A5D11D CRC64;

Query Match 29.4%; Score 58; DB 1; Length 875;

Best Local Similarity 30.6%; Pred. No. 29;

Matches 15; Conservative 5; Mismatches 15; Indels 14; Gaps 1;

QY 1 IEGPTLRQWLAARAGCGCGGGIE-----GPTLRQWLAAR 35

II IIII IIII IIII

Db 825 VQAHTLLAELLAPYGRGCGKKAISAQSSAELPQIEFLNKTLRQWISTQ 873

RESULT 5

ID HXD9_MOUSE STANDARD; PRT; 339 AA.
AC P28357;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Homeobox protein Hox-D9 (Hox-4.4) (Hox-5.2).
GN HOXD9 OR HOXD-9 OR HOX-4.4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92224884; PubMed=1348690;
RA Renucci A.G.P., Zappavigna V., Zakany J., Izpisua-Belmonte J.-C.,
RA Buerki K., Duboule D.;
RT "Comparison of mouse and human HOX-4 complexes defines conserved
RT sequences involved in the regulation of Hox-4.4.";
RL EMBO J. 11:1459-1468(1992).
RN [2]
RP SEQUENCE OF 272-331 FROM N.A.
RX MEDLINE=89356622; PubMed=2569970;
RA Dolle P., Duboule D.;
RT "Two gene members of the murine HOX-5 complex show regional and cell-
RT type specific expression in developing limbs and gonads.";
RL EMBO J. 8:1507-1515(1989).

-!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.

-!- SUBCELLULAR LOCATION: Nuclear.

-!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LIMB BUDS.

-!- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEBOX PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

EMBL; X62669; CAAM4542.1; -

EMBL; X14714; CAB57813.1; -

PIR; S09398; S09398.

PIR; S09569; S09569.

PIR; S20880; S20880.

HSSP; P02834; 1B81.

TRANSFAC; T01755; -

MGD; MGI:96210; Hoxd9.

InterPro; IPR001356; Homeobox.

Pfam; PF00046; homeobox; 1.

PRINTS; PR00024; HOMEBOX.

SMART; SM00389; HOX; 1.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-31

Perfect score: 197

Sequence: 1 IEPTLRQWLAARAGCGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63.5	32.2	209	1 NT5_RAT	P34131 rattus norv
2	60.5	30.7	210	1 NT5_HUMAN	P34130 homo sapien
3	60	30.5	497	1 FXD2_HUMAN	O60548 homo sapien
4	58	29.4	875	1 SYA_CHLTR	O84754 chlamydia t
5	57	28.9	339	1 HXD9_MOUSE	P28357 mus musculus
6	56.5	28.7	105	1 INS_BOVIN	P01317 bos taurus
7	56.5	28.7	105	1 INS_SHEEP	P01318 ovis aries
8	56	28.4	619	1 LAC1_NEUCR	P06811 neurospora
9	56	28.4	619	1 LAC2_NEUCR	P10574 neurospora
10	55	27.9	556	1 NUMB_DROME	P16554 drosophila
11	55	27.9	694	1 FZD8_HUMAN	Q9H461 homo sapien
12	55	27.9	875	1 SYA_CHLMU	Q9PLH5 chlamydia m
13	54.5	27.7	112	1 LSR2_MYCTU	O06285 mycobacteri
14	54.5	27.7	198	1 HS70_SCHJA	P12795 schistosoma
15	54.5	27.7	1028	1 OVO_DROME	P51521 drosophila
16	54	27.4	201	1 YR21_TRSVR	P25245 tomato ring
17	54	27.4	620	1 YR70_METJA	Q58280 methanococc
18	54	27.4	1001	1 ORK1_DROME	Q94526 drosophila
19	53.5	27.2	266	1 SCQ2_HUMAN	O43819 homo sapien
20	53.5	27.2	562	1 SVK_AERPE	O9YFI9 aspergillum p
21	53	26.9	497	1 GATA_MYCLE	O33105 mycobacteri
22	53	26.9	176	1 E2BE_RAT	Q43350 rattus norv
23	52.5	26.6	174	1 SSB_RHOSH	Q9ZAG8 rhodobacter
24	52.5	26.6	341	1 SPIN_CBEFV	P23061 choriostoneu
25	52.5	26.6	370	1 CVB_MICIK	Q9MLK2 micropechis
26	52.5	26.6	448	1 RU17_DROME	P17133 drosophila
27	52.5	26.6	969	1 PAC4_HUMAN	P29122 homo sapien
28	52	26.4	333	1 SIX3_MOUSE	Q62233 mus musculus
29	52	26.4	394	1 FXD3_CHICK	P79772 gallus gall
30	52	26.4	426	1 HKLB_LYCES	O22300 lycopersico
31	52	26.4	443	1 OC3N_HUMAN	P20265 homo sapien
32	52	26.4	445	1 OC3N_MOUSE	P31360 mus musculus
33	52	26.4	448	1 SRF_XENLA	P23790 xenopus lae

RESULT 1

ID	NT5_RAT	STANDARD	PRT	209 AA
AC	P34131			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Neurotrophin-5 precursor (NT-5) (Neutrophilic factor 5) (Neurotrophin-4)			
DE	(NT-4) (Neutrophilic factor 4).			
GN	NTF5 OR NTF4 OR NT4.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92212967; PubMed=1313578;			
RA	IP N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,			
RA	Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,			
RA	Yancopoulos G.D.;			
RT	"Mammalian neurotrophin-4: structure, chromosomal localization,			
RT	tissue distribution, and receptor specificity.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=92075279; PubMed=1742028;			
RA	Berkmeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,			
RA	Rosenthal A.;			
RT	"Neurotrophin-5: a novel neurotrophic factor that activates trk and			
RT	trkB.";			
RL	Neuron 7:857-866(1991).			
CC	-1- FUNCTION: COULD SERVE AS A TARGET-DERIVED TROPHIC FACTOR FOR			
CC	SENSORY AND SYMPATHETIC NEURONS.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, MUSCLE, OVARY, BRAIN,			
CC	HEART, STOMACH AND KIDNEY. EXPRESSED IN BOTH EMBRYO AND ADULT			
CC	TISSUES			
CC	-1- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; M86742; AAA41728.1; -			
DR	EMBL; S69323; AAB20548.1; -			
DR	PIR; JH0504; JH0504.			
DR	PIR; B42687; B42687.			
DR	HSSP; P34130; 1B8M.			
DR	InterPro; IPR002072; NGF.			
DR	Pfam; PF00243; NGF; 1.			
DR	PRINTS; PR00268; NGF.			
DR	ProDom; PD002052; NGF; 1.			
DR	SMART; SM00140; NGF; 1.			

ALIGNMENTS

008580 mus musculu
053258 mycobacteri
P20438 saccharomyc
P20437 saccharomyc
P1252 drosophila
P13645 homo sapien
P18681 neurospora
P38681 acanthamoeb
P93527 sorghum bic
O62376 mus musculu
P08621 homo sapien
Q98937 gallus gall

34 52 26.4 462 1 ERR1_MOUSE
35 52 26.4 494 1 GATA_MYCTU
36 52 26.4 545 1 CG12_YEAST
37 52 26.4 546 1 CG11_YEAST
38 52 26.4 584 1 CNAL_DROME
39 52 26.4 593 1 KICJ_HUMAN
40 52 26.4 1168 1 MYSC_ACACA
41 52 26.4 1176 1 NIR_NEUCR
42 52 26.4 1178 1 PHYB_SORBI
43 51.5 26.1 378 1 RU17_MOUSE
44 51.5 26.1 437 1 RU17_HUMAN
45 51.5 26.1 440 1 FXGA_CHICK

A:Residues: 1-163 <LAM>
 A:Cross-references: EMBL:AF067609; PIDN:AAC17537.1; GSPDB:GN00022; CESP:C23H5.9
 A:Experimental source: strain Bristol N2; clone C23H5
 C:Genetics:
 A:Gene: CESP:C23H5.9
 A:Map position: 4
 A:Introns: 1/3; 101/3; 126/2

Query Match 28.4%; Score 56; DB 2; Length 163;
 Best Local Similarity 75.0%; Pred. No. 22;

Matches 12; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 15 GCGCGGGGIEG--PTL 28

Db 33 GCGCGGGGCGCLPTL 48

RESULT 18

E86405
 hypothetical protein F13K9.10 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C:Accession: E86405
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E86405
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-349 <S1>
 A:Cross-references: GB:AE005172; NID:g11024859; PIDN:AAG26943.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 28.4%; Score 56; DB 2; Length 349;

Best Local Similarity 66.7%; Pred. No. 43;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 11 AARAGGCGGGGIEG 25

Db 333 ASCGGGCGGGCGG 347

RESULT 19

S41943
 cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete Phanerochaete chrysosporium
 C:Species: Phanerochaete chrysosporium
 C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C:Accession: S41943
 R:Sims, P.F.G.; Soares-Felipe, M.S.; Wang, Q.; Gent, M.E.; Tempelaars, C.; Broda, P.
 Mol. Microbiol. 12, 209-216, 1994
 A:Title: Differential expression of multiple exo-cellobiohydrolase I-like genes in the
 A:Reference number: S44714; MUID:94335641
 A:Accession: S44715
 A:Molecule type: mRNA
 A:Residues: 1-510 <S12>
 A:Cross-references: EMBL:Z29653; NID:g453222; PIDN:CAA82762.1; PID:g453224
 C:Genetics:
 A:Introns: 505/1
 C:Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homolog
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:479-510/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 28.4%; Score 56; DB 2; Length 510;

Best Local Similarity 48.0%; Pred. No. 59;

Matches 12; Conservative 1; Mismatches 6; Indels 6; Gaps 1;
 QY 3 GPTLROWLAARAGGCGGGGIEGPT 27
 Db 473 GPTVPQW-----GCGGGIGYSGST 491

RESULT 20

S44716
 cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chrysosporium)
 C:Species: Phanerochaete chrysosporium
 C:Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
 C:Accession: S44716; S33165
 R:Sims, P.F.G.; Soares-Felipe, M.S.; Wang, Q.; Gent, M.E.; Tempelaars, C.; Broda, P.
 Mol. Microbiol. 12, 209-216, 1994
 A:Title: Differential expression of multiple exo-cellobiohydrolase I-like genes in the
 A:Reference number: S44714; MUID:94335641
 A:Accession: S44716
 A:Molecule type: DNA
 A:Residues: 1-511 <S1M>
 A:Cross-references: EMBL:Z22527; NID:g296028; PIDN:CAA80252.1; PID:g3980202
 C:Genetics:
 A:Introns: 201/3; 506/1
 C:Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homolog
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:480-511/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 28.4%; Score 56; DB 2; Length 511;

Best Local Similarity 48.0%; Pred. No. 59;

Matches 12; Conservative 1; Mismatches 6; Indels 6; Gaps 1;

QY 3 GPTLROWLAARAGGCGGGGIEGPT 27

Db 474 GPTVPQW-----GCGGGIGYSGST 492

RESULT 21

S41942
 cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chrysosporium)
 C:Species: Phanerochaete chrysosporium
 C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C:Accession: S44714; S41942
 R:Sims, P.F.G.; Soares-Felipe, M.S.; Wang, Q.; Gent, M.E.; Tempelaars, C.; Broda, P.
 Mol. Microbiol. 12, 209-216, 1994
 A:Title: Differential expression of multiple exo-cellobiohydrolase I-like genes in the
 A:Reference number: S44714; MUID:94335641
 A:Accession: S44714
 A:Molecule type: mRNA
 A:Residues: 1-540 <S12>
 A:Cross-references: EMBL:Z29653; NID:g453222; PIDN:CAA82761.1; PID:g453223
 C:Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homolog
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:479-510/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 28.4%; Score 56; DB 2; Length 540;

Best Local Similarity 48.0%; Pred. No. 62;

Matches 12; Conservative 1; Mismatches 6; Indels 6; Gaps 1;

QY 3 GPTLROWLAARAGGCGGGGIEGPT 27

Db 473 GPTVPQW-----GCGGGIGYSGST 491

RESULT 22

KSNCL0
 laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)
 N:Alternate names: urishiol oxidase
 C:Species: Neurospora crassa
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
 C:Accession: A28523; A29762
 R:Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
 J. Biol. Chem. 263, 885-896, 1988
 A:Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino-

RESULT 16

T12783

C:Species: Bacillus subtilis phage SPBc2

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999

C:Accession: T12783; H69719

R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mael, C.; Karamata, D.

A:Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 protein

A:Reference number: Z17583

A:Accession: T12783

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-56 <LAZ>

A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025497; PIDN:AAC12992.1

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C:Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A:Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: H69719

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-56 <KUN>

A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14066.1; PID:el1833

A:Experimental source: strain 168

C:Genetics: <LAZ>

A:Gene: yolG

C:Genetics: <KUI>

A:Gene: sunA

C:Superfamily: unassigned lanthionine-containing peptides

C:Keywords: antibiotic; lanthionine

F:1-15/Domain: propeptide #status predicted <PRO>

F:20-56/Product: sublancin 168 #status predicted <MAT>

F:26-55/Disulfide bonds: #status experimental

F:33-48/Disulfide bonds: #status predicted

F:35/Modified site: dehydroalanine (Ser) #status experimental

F:38-41/cross-link: (2S,3S,6R)-3-methyl-lanthionine (Thr-Cys) #status predicted

Query Match 28.4%; Score 56; DB 2; Length 56;

Best Local Similarity 44.0%; Pred. No. 8.7;

Matches 11; Conservative 3; Mismatches 9; Indels 2; Gaps 1;

QY 9 WLAARAGG--GCGGGGEGPTLRQW 31

DB 30 WLQASGGTITGCGGAVACQNYRQF 54

RESULT 17

T33130

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33130

R:Lamar, E.; Kramer, J.

A:Description: The sequence of C. elegans cosmid C23H5.

A:Reference number: Z21286

A:Accession: T33130

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

RESULT 12

A-Molecule type: DNA

A-Status: nucleic acid

A:Molecule type: DNA

Quarry Match

037 3 CPTI.ROWI.-----AABACGGCGGGCTFCP 26

DECEMBER 1973

Biochem. J. 53, 366-374, 1953
A:Title: The amino-acid sequence in the glyceryl chain of insulin. 2. The investigation
A:Reference number: A90342
A:Accession: A90342
A:Molecule type: protein
A:Residues: 85-105 <SAN>
R:Sanger, F. Tuppy, H.
Biochem. J. 49, 481-490, 1951
A:Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investi
A:Reference number: A90341

C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8.09368 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838C-31

Perfect score: 197

Sequence: 1 EGGPTLROWLAARAGGCGGGIEGPTLROWLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	67	34.0	865	2 T34584	probable secreted
2	63.5	32.2	209	2 B42687	neurotrophin-4 pre
3	60.5	30.7	210	2 A42687	neurotrophin-4 pre
4	60	30.5	500	2 T20961	hypothetical prote
5	59.5	30.2	488	2 G87033	probable ATP/GTP-b
6	59.5	30.2	518	2 S72938	hflx protein - Myc
7	59	29.9	683	2 B71325	conserved hypothet
8	58	29.4	302	2 S71334	acetyl xylan ester
9	58	29.4	924	2 E71476	alanine--trNA liga
10	57.5	29.2	495	2 D70503	probable HflX - My
11	57	28.9	327	2 AG1974	hypothetical prote
12	57	28.9	339	2 S20880	homeotic protein H
13	56.5	28.7	77	1 INSH	insulin precursor
14	56.5	28.7	105	1 IFBO	insulin precursor
15	56.5	28.7	176	2 B72698	hypothetical prote
16	56	28.4	56	2 T12783	sublancin 168 prec
17	56	28.4	163	2 T33130	hypothetical prote
18	56	28.4	349	2 E86405	hypothetical prote
19	56	28.4	510	2 S41943	cellulose 1.4-beta
20	56	28.4	511	2 S44716	cellulose 1.4-beta
21	56	28.4	540	2 S41942	cellulose 1.4-beta
22	56	28.4	619	1 KSNCLT	laccase (EC 1.10.3
23	56	28.4	619	1 KSNCLT	laccase (EC 1.10.3
24	56	28.4	767	1 E70895	hypothetical glyci
25	55	27.9	180	2 T49530	related to glycine
26	55	27.9	201	2 T49792	hypothetical prote
27	55	27.9	257	2 C84890	hypothetical prote
28	55	27.9	331	2 T26807	hypothetical prote
29	55	27.9	333	2 T26808	hypothetical prote

RESULT 1

T34584

Probable secreted proteinase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 24-Nov-1999

C:Accession: T34584

R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, January 1998

A:Reference number: 221548

A:Accession: T34584

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-865 <MUR>

A:Cross-references: EMBL:AL021529; PIDN:CAA16449.1; GSPDB:GN00070; SCOEDB:SC10A5.17

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC10A5.17

Query Match 34.0% Score 67; DB 2; Length 865;

Best Local Similarity 66.7%; Pred No 5.9;

Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 9 WLAARAGGCGGGIEGP 26

Db 651 WLAACAAGCGGGGTNPP 668

RESULT 2

B42687

neurotrophin-4 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999

C:Accession: B42687; JH0504; JH0505

R:IP, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Belluscio,

Proc. Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992

A:Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distri

A:Reference number: A42687; MUID:92212967

A:Accession: B42687

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-209 <IPA>

A:Cross-references: GB:M86742; NID:g205775; PIDN:AAA1728.1; PID:g205776

R:Berkemeier, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosenth

Neuron 7, 857-866, 1991

A:Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.

A:Reference number: JH0503; MUID:92075279

A:Accession: JH0504

A:Molecule type: DNA

A:Residues: 1-209 <BER>

A:Accession: JH0505

A:Molecule type: mRNA

A:Residues: 1-176, 'p', 178-209 <BER1>


```

; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 37.1%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 29
US-09-516-704-18
; Sequence 18, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprence, Randolph B.
; Podduturi, Surekha
; Deprence, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product="Beta-ala"
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```

```

US-09-516-704-18

Query Match 37.1%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 30
US-09-516-704-194
; Sequence 194, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwiria, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprence, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-09-516-704-194

Query Match 37.1%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

Search completed: October 9, 2002, 09:06:34
Job time : 6.98595 secs

```

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
US-09-244-298A-18

Query Match 37.18; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTTLROWLAARA 14
|||||
DB 1 IEPTTLROWLAARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 37.18; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTTLROWLAARA 14
|||||
DB 2 IEPTTLROWLAARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

;
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match 37.1%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 24
US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194

Query Match 37.1%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPTLRQWLAARA 15

RESULT 25
US-08-973-225-220
; Sequence 220, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match 37.1%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPTLRQWLAARA 15

RESULT 26
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.


```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-194
;
; Query Match 37.1%; Score 73; DB 2; Length 16;
; Best Local Similarity 100.0%; Pred. No. 0.016;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEPTLROWLAARA 14
; DB 2 IEPTLROWLAARA 15
;
; RESULT 23
; US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherril S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
; US-08-764-640-232
; Query Match 37.1%; Score 73; DB 2; Length 16;
; Best Local Similarity 100.0%; Pred. No. 0.016;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEPTLROWLAARA 14
; DB 2 IEPTLROWLAARA 15
;
; RESULT 23
; US-08-973-225-18
; Sequence 18, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherril S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
; US-08-764-640-232
; Query Match 37.1%; Score 73; DB 2; Length 16;
; Best Local Similarity 100.0%; Pred. No. 0.016;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEPTLROWLAARA 14
; DB 2 IEPTLROWLAARA 15
;
; RESULT 23
; US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherril S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

Db 1 IEGPTLRQWLAARA 14

RESULT 19

US-09-516-704-185
; Sequence 185, Application US/09516704
; Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprience, Randolph B.
; Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

MOLECULE TYPE: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-09-516-704-185

Query Match 37.1%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14

Db 2 IEGPTLRQWLAARA 15

RESULT 20

US-09-764-640-18
; Sequence 18, Application US/08764640
; Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.

APPLICANT: Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprience, Randolph B.
; Podduturi, Surekha
; Yia, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product="Beta-ala"

US-08-764-640-18

Query Match 37.1%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14

Db 1 IEGPTLRQWLAARA 14

RESULT 21

US-08-764-640-194
; Sequence 194, Application US/08764640
; Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprience, Randolph B.

APPLICANT: Podduturi, Surekha

APPLICANT: Yia, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

Wed Oct 9 10:30:09 2002

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17

Query Match 37.1%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015; 0; Indels
Matches 14; Conservative 0; Mismatches 0; Gaps 0;

QY 1 IEGPTLROWLAARA 14
|||||
Db 1 IEGPTLROWLAARA 14
|||||

RESULT 17
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-185

Query Match 37.1%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARA 14
|||||
Db 2 IEGPTLROWLAARA 15
|||||

RESULT 18
US-09-516-704-17
Sequence 17, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: RECEPTOR
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 37.1%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARA 14
|||||

RESULT 14
US-08-973-225-17
; Sequence 17, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17

Query Match 37.1%, Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14

RESULT 15
US-08-973-225-185
; Sequence 185, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.

; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185

Query Match 37.1%, Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEGPTLRQWLAAARA 14
Db 2 IEGPTLRQWLAAARA 15

RESULT 16
US-09-244-298A-17
; Sequence 17, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Poduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk

us-09-422-838c-31.ra1

Wed Oct 9 10:30:09 2002

```

; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hrubiec, Robert T.
;   REGISTRATION NUMBER: 36,392
;   REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 14 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: <Unknown>
;     TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match      37.1%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 1 IEGPTLQWLAAARA 14

RESULT 12
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 15 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-193

Query Match      37.1%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 1 IEGPTLQWLAAARA 14

RESULT 13
US-08-764-640-185
; Sequence 185, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 15 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-764-640-185

Query Match      37.1%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 2 IEGPTLQWLAAARA 15
```

us-09-422-838c-31.ra1

Wed Oct 9 10:30:09 2002

```

; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hrubiec, Robert T.
;   REGISTRATION NUMBER: 36,392
;   REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 14 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: <Unknown>
;     TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match      37.1%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 1 IEGPTLQWLAAARA 14

RESULT 12
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 15 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-193

Query Match      37.1%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 1 IEGPTLQWLAAARA 14

RESULT 13
US-08-764-640-185
; Sequence 185, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 15 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-764-640-185

Query Match      37.1%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 2 IEGPTLQWLAAARA 15
```

Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-193

Query Match 37.1%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14
|||||

RESULT 10
US-09-516-704-13
Sequence 13, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13

Query Match 37.1%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14
|||||

RESULT 11
US-09-516-704-193
Sequence 193, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

us-09-422-838c-31.ra1

Wed Oct 9 10:30:09 2002

```
US-08-973-225-193
;
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13
;
; Query Match 37.1%; Score 73; DB 3; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQWLAAARA 14
; Db 1 IEGPTLRQWLAAARA 14
;
; RESULT 8
; US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Poddaturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-Dec-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-13
;
; Query Match 37.1%; Score 73; DB 3; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQWLAAARA 14
; Db 1 IEGPTLRQWLAAARA 14
;
; RESULT 9
; US-09-244-298A-193
; Sequence 193, Application US/09244298A
;
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-08-973-225-193
;
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13
;
; Query Match 37.1%; Score 73; DB 3; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQWLAAARA 14
; Db 1 IEGPTLRQWLAAARA 14
;
; RESULT 7
; US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherrill S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
```

```

; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-13

Query Match 37.1%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 1 IEPTLRQWLAARA 14

RESULT 5
US-08-764-640-193
; Sequence 193, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR

```

```

; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-193

Query Match 37.1%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 1 IEPTLRQWLAARA 14

RESULT 6
US-08-973-225-13
; Sequence 13, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherril S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A

```



```

; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /product= "Ava"
US-08-764-640-231

```

Query Match 38.8%; Score 76.5; DB 2; Length 25;
Best Local Similarity 40.6%;
Matches 13; Conservative 8; Mismatches 2; Indels
Pred. No. 0.0098;

QY 2 EGPTLRQWLAAARAGCGGGGIEGPTLRQWLA 33
:|||||: :| :|||||: :
Db 2 DGPTLRWISFXA-----DGPTLRWIS 24

RESULT 2
US-09-244-298A-231
; Sequence 231, Application US/09244298A

APPLICANT: Dower, William J.
 APPLICANT: Barrett, Ronald W.
 APPLICANT: Cwiria, Steven E.
 APPLICANT: Gates, Christian
 APPLICANT: Schatz, Peter J.
 APPLICANT: Balasubramanian, Palaniappan
 APPLICANT: Wagstrom, Christopher R.
 APPLICANT: Hendren, Richard W.
 APPLICANT: Deprince, Randolph B.
 APPLICANT: Podduturi, Surekha
 APPLICANT: Yin, Qun

APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, V
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/05/244,298A
; FILING DATE: 11-DEC-1996
;

```

```

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrublec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
JS-09-244-298A-231

```

Query Match 38.8%; Score 76.5;
Best Local Similarity 40.6%; Pred. No. 0
Matches 13: Conservative 8, Mismatch 5

	Matches	13;	Conservative	8;	Mismatches	2;	Indels	9;	Gaps	1;
Dy	2	EGPTLRQLAARAGGGCGGGGIEGPTLRQLA	33							

100

100

```
Db      :|||||:|:|:|
2 DGPTLREWISFXA-----DGPTLREWIS 24
```

RESULT 3
US-09-516-704-231
; Sequence 231, Application US/09516704
; Patent No. 6251864

GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwila, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.

Podduturi, Surekha
 TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
 RECEPTOR
 NUMBER OF SEQUENCES: 244
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Glaxo Wellcome
 STREET: Five Moore Drive, P.O. Box 13398
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Ver 1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids

```

: TYPE: amino acid
: STRANDEDNESS: <Unknown>
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FEATURE:
:   NAME/KEY: Modified-site
:   LOCATION: 13
:   OTHER INFORMATION: /product="Ava"
:   SEQUENCE DESCRIPTION: SEQ ID NO: 231:
:   US-09-516-704-231

```

Query Match	38 89	Score
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
32	1	1
33	1	1
34	1	1
35	1	1
36	1	1
37	1	1
38	1	1
39	1	1
40	1	1
41	1	1
42	1	1
43	1	1
44	1	1
45	1	1
46	1	1
47	1	1
48	1	1
49	1	1
50	1	1
51	1	1
52	1	1
53	1	1
54	1	1
55	1	1
56	1	1
57	1	1
58	1	1
59	1	1
60	1	1
61	1	1
62	1	1
63	1	1
64	1	1
65	1	1
66	1	1
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	1
75	1	1
76	1	1
77	1	1
78	1	1
79	1	1
80	1	1
81	1	1
82	1	1
83	1	1
84	1	1
85	1	1
86	1	1
87	1	1
88	1	1
89	1	1
90	1	1
91	1	1
92	1	1
93	1	1
94	1	1
95	1	1
96	1	1
97	1	1
98	1	1
99	1	1
100	1	1

```

Best Local Similarity 40.8%; Score 78.3; DB 4; Length 25;
Matches 13; Conservative 8; Mismatches 2; Indels 9;

0y 2 EGPTRQWLAARAGCGGGGEGPTLRQWLA 33
    :||||:|:| :|:|||||:|:|
0b 2 DGPTLRWISFEXA-----DGPTLRWIS 24

RESULT 4
JS-08-764-640-13

```

Sequence 13, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION.

Wed Oct 0 10:30:09 2002

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 : Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-31
Perfect score: 197
Sequence: 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.5	38.8	25	2	US-08-764-640-231
2	76.5	38.8	25	3	US-09-244-298A-231
3	76.5	38.8	25	4	US-09-516-704-231
4	73	37.1	14	2	US-08-764-640-13
5	73	37.1	14	2	US-08-764-640-193
6	73	37.1	14	3	US-08-973-225-13
7	73	37.1	14	3	US-08-973-225-193
8	73	37.1	14	3	US-09-244-298A-13
9	73	37.1	14	4	US-09-244-298A-193
10	73	37.1	14	4	US-09-516-704-13
11	73	37.1	14	4	US-09-516-704-193
12	73	37.1	15	2	US-08-764-640-17
13	73	37.1	15	2	US-08-764-640-185
14	73	37.1	15	3	US-08-973-225-17
15	73	37.1	15	3	US-08-973-225-185
16	73	37.1	15	3	US-09-244-298A-17
17	73	37.1	15	3	US-09-244-298A-185
18	73	37.1	15	4	US-09-516-704-17
19	73	37.1	15	4	US-09-516-704-185
20	73	37.1	16	2	US-08-764-640-18
21	73	37.1	16	2	US-08-764-640-194
22	73	37.1	16	2	US-08-764-640-232
23	73	37.1	16	3	US-08-973-225-18
24	73	37.1	16	3	US-08-973-225-194
25	73	37.1	16	3	US-08-973-225-220
26	73	37.1	16	3	US-09-244-298A-18
27	73	37.1	16	3	US-09-244-298A-194

28	73	37.1	16	3	US-09-244-298A-232	Sequence 232, App
29	73	37.1	16	4	US-09-516-704-18	Sequence 18, Appl
30	73	37.1	16	4	US-09-516-704-194	Sequence 194, App
31	73	37.1	16	4	US-09-516-704-232	Sequence 232, App
32	69	35.0	14	2	US-08-764-640-195	Sequence 195, App
33	69	35.0	14	2	US-08-764-640-199	Sequence 199, App
34	69	35.0	14	3	US-08-973-225-195	Sequence 195, App
35	69	35.0	14	3	US-08-973-225-199	Sequence 199, App
36	69	35.0	14	3	US-09-244-298A-195	Sequence 195, App
37	69	35.0	14	3	US-09-244-298A-199	Sequence 199, App
38	69	35.0	14	4	US-09-516-704-195	Sequence 195, App
39	69	35.0	14	4	US-09-516-704-199	Sequence 199, App
40	69	35.0	15	2	US-08-764-640-196	Sequence 196, App
41	69	35.0	15	2	US-08-764-640-200	Sequence 200, App
42	69	35.0	15	2	US-08-764-640-209	Sequence 209, App
43	69	35.0	15	2	US-08-764-640-215	Sequence 215, App
44	69	35.0	15	3	US-08-973-225-196	Sequence 196, App
45	69	35.0	15	3	US-08-973-225-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-08-764-640-231
Sequence 231, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagsstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:

Qy	1	IEPTLRQWLAA	RAGGGGGG	GGGIEGPTLRQWL	AARA	36
Db	1	IEPTLRQCLAA	RAGGGGGG	GGGIEGPTLRQCL	AARA	36

RESULT 30
AAB17299
ID AAB17299 standard; Peptide; 36 AA.
XX
XX
XX AAB17299;
XX
XX 31-OCT-2000 (first entry)
XX
XX TPO-mimetic peptide sequence SEQ ID NO:355.
DE

Modified peptide; therapeutic agent; fusion; FC domain; cancer; autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CTU4A; mimetic; IL-1; TNF; antagonist; MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; vascular endothelial growth factor; matrix metalloproteinase; asthma; thrombosis; pharmaceutical.

Synthetic.

WO200024782-A2.

04-MAY-2000.

25-OCT-1999;

23-OCT-1998: 98US-0705371

44 OCT 1999; 9905-0428082.

(AMGEN) AMGEN INC.

rege U, Liu C, Cheetham J, Boone TC;

NET; 2000-350/02/30.

over composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

Example 1; Page 320-321; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-Fl-(X2)b, where: Fl = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-Pl, -(L1)c-P1-(L2)d-P2, (L1)c-Pl-(L2)d-P2-(L3)e-P3, or -(L1)c-Pl-(L2)d-P2-(L3)e-P3-(L4)f-P4 where Fl, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 1 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antitasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein binding, complement fixation, and possibly placental transfer. AAA69443 o AAB169526 and AAB169503 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

frequency 36 AA;

Query Match	80.7%;	Score 159;	DB 21;	Length 36;
Best Local Similarity	91.7%;	Pred. No. 6.1e-12;		
Matches	31.	Concave		

1 IEPTLRQWLARAGCGGGIEPTLRQWLAARA 36

dbb
1 IEGPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36

Search completed: October 9, 2002, 08:58:58
Job time : 17.1874 secs

Wed Oct '9 10:30:08 2002

Best Local Similarity 91.7%; Pred. No. 1.3e-12; Mismatches 0; Indels 3; Gaps 1;

Query Match 86.8%; Score 171; DB 21; Length 34;
Best Local Similarity 94.4%; Pred. No. 2.3e-13; Mismatches 0; Indels 2; Gaps 1;

QY 1 IEGPTLRQWLAAAGCGGGGIEGPTLRQWLAAARA 36
|||||
DB 1 IEGPTLRQWLAAAGCGGGGIEGPTLRQWLAAARA 34
|||||

RESULT 28

AAB17290
ID AAB17290 standard; Peptide; 33 AA.

AC AAB17290;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:346.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases.

XX Example 1; Page 317; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where F1, P2, P3, and P4 = are each independently sequences of
CC -(L1)c-P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 33 AA;

XX Query Match 83.5%; Score 164.5; DB 21; Length 33;

Best Local Similarity 91.7%; Pred. No. 1.3e-12; Mismatches 0; Indels 3; Gaps 1;

QY 1 IEGPTLRQWLAAAGCGGGGIEGPTLRQWLAAARA 36
|||||
DB 1 IEGPTLRQWLAAAGCGGGGIEGPTLRQWLAAARA 33
|||||

RESULT 29

AAB17298
ID AAB17298 standard; Peptide; 36 AA.

XX AAB17298;

XX 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:354.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases.

XX Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where F1, P2, P3, and P4 = are each independently sequences of
CC -(L1)c-P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

XX Query Match 80.7%; Score 159; DB 21; Length 36;

XX Best Local Similarity 91.7%; Pred. No. 6.1e-12; Mismatches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 39 AA;
Query Match 88.1%; Score 173.5; DB 21; Length 39;
Best Local Similarity 89.7%; Pred. No. 1.3e-13;
Matches 35; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 1 IEGPTLROWLAARAGG---CGGGGIEGPTLROWLAARA 36
|||||
Db 1 IEGPTLROWLAARAGGKPEGGGGIEGPTLROWLAARA 39
|||||

RESULT 26

AAB17296

ID AAB17296 standard; Peptide; 42 AA.

XX AC AAB17296;

XX DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:352.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX where P1, P2, P3, and P4 = are each independently sequences of

XX pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX independently linkers; and a, b, c, d, e, and f = are each independently

XX 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX activities. DNAs, vectors and host cells from the present invention can

XX be used for producing pharmaceutical compositions. The compositions are

XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

XX The use of an Fc domain (rather than a Fab domain) can provide a longer

XX half-life or incorporate functions such as Fc receptor binding, protein

XX A binding, complement fixation, and possibly placental transfer. AAA69443

XX to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

XX sequences used in the exemplification of the present invention.

XX
SQ Sequence 42 AA;

Query Match 87.3%; Score 172; DB 21; Length 42;

Best Local Similarity 83.3%; Pred. No. 2.1e-13;

Matches 35; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 IEGPTLROWLAARA-----GGGGGIEGPTLROWLAARA 36

|||||

Db 1 IEGPTLROWLAARAGGGGGGGGGGIEGPTLROWLAARA 42

|||||

RESULT 27

AAB17291

ID AAB17291 standard; Peptide; 34 AA.

XX AC AAB17291;

XX DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:347.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases -

XX Example 1; Page 317; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX where P1, P2, P3, and P4 = are each independently sequences of

XX pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX independently linkers; and a, b, c, d, e, and f = are each independently

XX 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX activities. DNAs, vectors and host cells from the present invention can

XX be used for producing pharmaceutical compositions. The compositions are

XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

XX The use of an Fc domain (rather than a Fab domain) can provide a longer

XX half-life or incorporate functions such as Fc receptor binding, protein

XX A binding, complement fixation, and possibly placental transfer. AAA69443

XX to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

XX sequences used in the exemplification of the present invention.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC
half

-life or incorporate functions such as Fc

receptor
acceptal

binding, protein transfer AAA694A3

CC or E; X₉ = W, Y or F; X₁₋₁₀ = L, I, V, A, F, M, or K; X₁₋₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₋₁₂ = A, I, V, L, F, G, S, or Q; X₁₋₁₃ = R, K,
 CC T, V, N, Q or G; X₁₋₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The RMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 36 AA;

Query Match 90.9%; Score 179; DB 21; Length 36;

Best Local Similarity 94.4%; Pred. No. 2.8e-14;

Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARAGCGGGGIEGPTLROWLAARA 36

DB 1 IEGPTLROWLAARAGCGGGGIEGPTLROWLAARA 36

RESULT 22

AAB17292

ID AAB17292 standard; Peptide; 35 AA.

XX AC AAB17292;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:348.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

XX PT pharmacologically active peptides, useful for treating cancer and

XX PT autoimmune diseases -

XX PS Example 1; Page 317-318; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

XX CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX CC where P1, P2, P3, and P4 = are each independently sequences of

XX CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX CC independently linkers; and a, b, c, d, e, and f = are each

XX CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX CC activities. DNAs, vectors and host cells from the present invention can

XX CC be used for producing pharmaceutical compositions. The compositions are

CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 35 AA;

Query Match 90.1%; Score 177.5; DB 21; Length 35;

Best Local Similarity 97.2%; Pred. No. 4.1e-14;

Matches 35; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 IEGPTLROWLAARAGCGGGGIEGPTLROWLAARA 36

DB 1 IEGPTLROWLAARAGGG-GGGGIEGPTLROWLAARA 35

RESULT 23

AAB17294

ID AAB17294 standard; Peptide; 37 AA.

XX AC AAB17294;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:350.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

XX PT pharmacologically active peptides, useful for treating cancer and

XX PT autoimmune diseases -

XX PS Example 1; Page 318; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

XX CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX CC where P1, P2, P3, and P4 = are each independently sequences of

XX CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX CC independently linkers; and a, b, c, d, e, and f = are each

XX CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX CC activities. DNAs, vectors and host cells from the present invention can

XX CC be used for producing pharmaceutical compositions. The compositions are

0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Query Match 91.9%; Score 181; DB 21; Length 268;
Best Local Similarity 97.1%; Pred. No. 1.2e-13;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGCGGGGEGPTLRQWLAAAR 35
DB 234 IEGPTLRQWLAAARAGCGGGGEGPTLRQWLAAAR 268

RESULT 20
AAB17306
ID AAB17306 standard; Peptide; 36 AA.
XX AC AAB17306;
XX DT 31-OCT-2000 (first entry)
XX DE TPO-mimetic peptide sequence SEQ ID NO:362.
XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX OS Synthetic.
XX PN WO200024782-A2.
XX PF 04-MAY-2000.
XX PP 25-OCT-1999; 99WO-US25044.
XX PR 23-OCT-1998; 98US-0105371.
XX PR 22-OCT-1999; 99US-042082.
XX PA (AMGE-) AMGEN INC.
XX PI Felge U, Liu C, Cheetham J, Boone TC;
XX DR WPI; 2000-350702/30.
XX PT Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -
XX PS Example 1; Page 324; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. CC The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein CC A binding, complement fixation, and possibly placental transfer. AAA69443 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid CC sequences used in the exemplification of the present invention.

XX SQ Sequence 36 AA;
Query Match 90.9%; Score 179; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 2.8e-14;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGCGGGGEGPTLRQWLAAAR 36
DB 1 IEGPTLRQWLAAARAGCGGGGEGPTLRQWLAAAR 36

RESULT 21
AAY96526
ID AAY96526 standard; peptide; 36 AA.
XX AC AAY96526;
XX DT 04-SEP-2000 (first entry)
XX DE Thrombopoietin mimetic peptide compound 7.
XX KW Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX OS Synthetic.
XX FH Key
FT Modified-site 1 Location/Qualifiers
FT Peptide 1..14 /note= "optionally linked to an Fc molecule"
FT Peptide 15..18 /label= TMP_1
FT Peptide 19..32 /label= linker
FT Peptide /label= TMP_2
XX PN WO200024770-A2.
XX PP 04-MAY-2000.
XX PF 22-OCT-1999; 99WO-US24834.
XX PR 23-OCT-1998; 98US-0105348.
XX PA (AMGE-) AMGEN INC.
XX PI Liu C, Felge U, Cheetham J;
XX DR WPI; 2000-365108/31.
XX PT Thrombopoietic peptides which activate mpl receptors and increase the
XX production of platelets or platelet precursors, useful for treatment of
XX diseases which involve thrombocytopenia
XX PS Claim 16; Page 62; 91pp; English.

A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2], is new. TMP-1 and TMP-2 are amino acid sequences varying from at least 10 to 14 residues in length comprising X2-X1.0, X2-X1.1, X2-X1.2, X2-X1.3, X2-X1.4, X1-X1.0, X1-X1.1, X1-X1.2, X1-X1.3, and X1-X1.4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A; X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,

CC Fc domain, pharmacologically active peptides, and linkers. Where (1) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently selected sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, or
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA;

Query Match 93.9%; Score 185; DB 21; Length 269;
 Best Local Similarity 97.2%; Pred. No. 4.1e-14;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARAGCGGGGIEGPTLQWLAAARA 36
 DB 2 IEGPTLQWLAAARAGCGGGGIEGPTLQWLAAARA 37

RESULT 18

AAAY96531
 ID AAY96531 standard; Protein; 269 AA.

AC AAY96531;

DT 04-SEP-2000 (first entry)

DE Human IgG1 Fc Tmp fusion protein.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.

XX Homo sapiens.

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-0524834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX N-PSDB; AAA29229.

PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Example 2A; Page 49-50; 91pp; English.

CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP-1-(L1)-TMP-2),
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X2-X1-0, X2-X1-1, X2-X1-2,
 CC X2-X1-3, X2-X1-4, X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and
 CC X1-X1-4, X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
 CC X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,

CC or E; X9 = W, Y or F; X1-0 = L, I, V, A, F, M, or K; X1-1 = A, I, V,
 CC L, F, S, T, K, H, or E; X1-2 = A, I, V, L, F, G, S, or Q; X1-3 = R, K,
 CC T, V, N, Q or G; X1-4 = A, I, V, L, F, T, R, E, or G; L1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The Tmps are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 269 AA;

Query Match 93.9%; Score 185; DB 21; Length 269;
 Best Local Similarity 97.2%; Pred. No. 4.1e-14;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARAGCGGGGIEGPTLQWLAAARA 36
 DB 234 IEGPTLQWLAAARAGCGGGGIEGPTLQWLAAARA 269

RESULT 19

AAB16959
 ID AAB16959 standard; Protein; 268 AA.

XX AC AAB16959;

XX 31-OCT-2000 (first entry)

XX Fc-TMP protein sequence SEQ ID NO:8.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX N-PSDB; AAA69445.

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 182-183; 608pp; English.

CC The present invention describes composition of matter (1) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (1) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently selected sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently selected linkers; and a, b, c, d, e, and f = are each independently

CC was then fused in-frame to the Fc region of the human IgG1 chain (see
 CC AA96529). A compound which binds to an mPl receptor comprising a TMP
 CC dimer joined by a linker [TMP-1-(L1)-nTMP-2], is new. TMP-1 and TMP-2
 CC are amino acid sequences varying from at least 10 to 14 residues in
 CC length comprising X₂-X₁-X₀, X₂-X₁-X₁, X₂-X₁-X₂, X₂-X₁-X₃, X₂-X₁-X₄,
 CC X₁-X₁-X₀, X₁-X₁-X₁, X₁-X₁-X₂, X₁-X₁-X₃, and X₁-X₁-X₄. X₁ = I, A,
 CC V, L, S or R; X₂ = E, D, K or V; X₃ = G or A; X₄ = P; X₅ = T or S;
 CC X₆ = L, I, V, A, F, M, or K; X₇ = R or K; X₈ = Q, N, or E; X₉ = W, Y or F;
 CC X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V, L, F, S, T, K, H, or E;
 CC X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K, T, V, N, Q or G; X₁₄ =
 CC A, I, V, L, F, T, R, E, or G; L₁ = linker comprising 1 to 20 amino
 CC acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl
 CC receptor which mediates the activity of endogenous thrombopoietin. The
 CC TMPs are useful for increasing the production of platelets or platelet
 CC precursors (e.g. megakaryocytes) in a mammal, which is useful for
 CC treatment of diseases which involve thrombocytopenia, e.g. aplastic
 CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
 CC associated ITP, and systemic lupus erythematosus.

XX Sequence 42 AA;

Query Match 93.9%; Score 185; DB 21; Length 42;
 Best Local Similarity 97.2%; Pred. No. 6.5e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 ||||||||||||||| |||||||||||||||
 Db 7 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 16

AAAB17311
 ID AAB17311 standard; Peptide; 60 AA.

AC AAB17311;

DT 31-OCT-2000 (first entry)

DE Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.
 OS Synthetic.

XX WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PF 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 331; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement function, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 60 AA;

Query Match 93.9%; Score 185; DB 21; Length 60;
 Best Local Similarity 97.2%; Pred. No. 9.2e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

||||||||||||| |||||||||||||||
 Db 2 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37

RESULT 17

AAB16960

ID AAB16960 standard; Protein; 269 AA.

AC AAB16960;

DT 31-OCT-2000 (first entry)

DE TMP-TMP-Fc protein sequence SEQ ID NO:10.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.
 OS Synthetic.

XX WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PF 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX N-PSDB; AAA69446.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 185-186; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

PT pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 PS Disclosure; Page 313; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P³, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 42 AA;

Query Match 93.9%; Score 185; DB 21; Length 42;
 Best Local Similarity 97.2%; Pred. No. 6.5e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 ||||||||||||||| |||||||||||||||
 DB 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 14
 AAB17308
 ID AAB17308 standard; Peptide; 42 AA.
 XX
 AC AAB17308;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Homo sapiens.
 OS
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 DR WPT; 2000-350702/30.
 XX
 XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -
 XX
 PS Example 2; Page 327; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P³, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 42 AA;

Query Match 93.9%; Score 185; DB 21; Length 42;
 Best Local Similarity 97.2%; Pred. No. 6.5e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 ||||||||||||||| |||||||||||||||
 DB 7 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 15
 AAY96530
 ID AAY96530 standard; Protein; 42 AA.
 XX
 AC AAY96530;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Thrombopoietin mimetic peptide.
 XX
 KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.
 XX
 OS Synthetic.
 XX
 PN WO200024770-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX
 PR 23-OCT-1998; 98US-0105348.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Liu C, Feige U, Cheetham J;
 XX
 DR WPI; 2000-365108/31.
 DR N-PSDB; AAA29225.
 XX
 XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Example 2A; Page 48; 91pp; English.
 XX
 CC Overlapping oligonucleotides were used to construct a synthetic
 CC gene encoding a thrombopoietin mimetic peptide (TMP), which

PT The present invention describes composition of matter (I) comprising an
FT Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 36 AA;
Query Match 93.9%; Score 185; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 5.6e-15;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 IEPTTLRWLAARAGGGGGGIEGPTLRQWLAARA 36
Db 1 IEPTTLRWLAARAGGGGGGIEGPTLRQWLAARA 36
RESULT 8
AA96523
ID AAY96523 standard; peptide; 36 AA.
AC AAY96523;
XX
XX 04-SEP-2000 (first entry)
XX Thrombopoietin mimetic peptide compound 4.
XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "optionally linked to an Fc molecule"
FT Peptide 1..14 /label= TMP_1
FT Peptide 15..22 /label= linker
FT Modified-site 18 /note= "optionally modified by bromoacetyl or peg"
FT Peptide 23..36 /label= TMP_2
XX WO200024770-A2.
XX
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-US24834.
XX 23-OCT-1998; 98US-0105348.
XX (AMGE-) AMGEN INC.
XX Liu C, Feige U, Cheetham J;
XX WPI; 2000-365108/31.
XX

PT Thrombopoietic peptides which activate mpl receptors and increase the
FT production of platelets or platelet precursors, useful for treatment of
XX diseases which involve thrombocytopenia
XX Claim 16; Page 62; 9lpp; English.
XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker (TMP_1-(L1)-TMP_2),
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X2-X1-0, X2-X1-1, X2-X1-2,
CC X2-X1-3, X2-X1-4, X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and
CC X1-X1-4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
CC X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
CC or E; X9 = W, Y or F; X10 = L, I, V, A, F, M, or K; X11 = A, I, V,
CC L, F, S, T, K, H, or E; X12 = A, I, V, L, F, G, S, or Q; X13 = R, K,
CC T, V, N, Q or G; X14 = A, I, V, L, F, T, E, or G; L1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 36 AA;
Query Match 93.9%; Score 185; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 5.6e-15;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 IEPTTLRWLAARAGGGGGGIEGPTLRQWLAARA 36
Db 1 IEPTTLRWLAARAGGGGGGIEGPTLRQWLAARA 36
RESULT 9
AA96525
ID AAY96525 standard; peptide; 36 AA.
XX
XX AAY96525;
XX
XX 04-SEP-2000 (first entry)
XX Thrombopoietin mimetic peptide compound 6.
XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "optionally linked to an Fc molecule"
FT Peptide 1..14 /label= TMP_1
FT Peptide 15..18 /label= linker
FT Peptide 19..32 /label= TMP_2
FT Modified-site 32 /note= "optionally linked to an Fc molecule"
XX WO200024770-A2.
XX
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-US24834.
XX 23-OCT-1998; 98US-0105348.
XX (AMGE-) AMGEN INC.
XX Liu C, Feige U, Cheetham J;
XX WPI; 2000-365108/31.
XX

PT pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 PS Disclosure; Page 190; 608pp; English.
 XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX Sequence 36 AA;

Query Match 93.9%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 5.6e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36
 |||||
 Db 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36

RESULT 6
 AAB17293
 ID AAB17293 standard; Peptide; 36 AA.

XX AAB17293;
 XX 31-OCT-2000 (first entry)
 XX TPO-mimetic peptide sequence SEQ ID NO:349.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 XX WO200024782-A2.
 XX 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX
 PS

Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 XX Sequence 36 AA;

Query Match 93.9%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 5.6e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36
 |||||
 Db 1 IEGPTLRQWLAAARAGGGGGGEGPTLRQWLAAARA 36

RESULT 7
 AAB17301
 ID AAB17301 standard; Peptide; 36 AA.

XX AAB17301;
 XX 31-OCT-2000 (first entry)
 XX TPO-mimetic peptide sequence SEQ ID NO:357.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 XX WO200024782-A2.
 XX 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

Example 1; Page 321; 608pp; English.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 16.1874 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-31

Perfect score: 197
Sequence: 1 IEPTLRQLARAGCGGGIEPTLRQLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseg/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197	100.0	36	21	AA17303
2	197	100.0	36	21	AA17307
3	197	100.0	36	21	AA17304
4	185.5	94.2	39	21	AA17305
5	185	93.9	36	21	AA16963
6	185	93.9	36	21	AA17293
7	185	93.9	36	21	AA17301
8	185	93.9	36	21	AA17302
9	185	93.9	36	21	AA17302
10	185	93.9	40	21	AA17302
11	185	93.9	41	21	AA196528

12	185	93.9	42	21	AA17281
13	185	93.9	42	21	AA17282
14	185	93.9	42	21	AA17308
15	185	93.9	42	21	AA196530
16	185	93.9	60	21	AA17311
17	185	93.9	269	21	AA16960
18	185	93.9	269	21	AA196531
19	181	91.9	268	21	AA16959
20	179	90.9	36	21	AA17306
21	179	90.9	36	21	AA196526
22	177.5	90.1	35	21	AA17292
23	174.5	88.6	37	21	AA17294
24	174	88.3	38	21	AA17295
25	173.5	88.1	39	21	AA17304
26	172	87.3	42	21	AA17296
27	171	86.8	34	21	AA17291
28	164.5	83.5	33	21	AA17290
29	159	80.7	36	21	AA17298
30	159	80.7	36	21	AA17299
31	159	80.7	36	21	AA196521
32	158	80.2	32	21	AA17289
33	157	79.7	36	21	AA17300
34	157	79.7	36	21	AA196522
35	151.5	76.9	31	21	AA17288
36	145	73.6	30	21	AA17287
37	144	73.1	32	21	AA17297
38	144	73.1	32	21	AA196520
39	144	73.1	34	21	AA196527
40	138.5	70.3	29	21	AA17286
41	132	67.0	28	21	AA17285
42	131.5	66.8	29	21	AA16970
43	129.5	65.7	31	21	AA16973
44	129.5	65.7	31	21	AA16974
45	125.5	63.7	29	21	AA16971

ALIGNMENTS

RESULT 1
AA17303
ID AA17303 standard; Peptide; 36 AA.
XX AA17303;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:359.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antitumour; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

DR WPI; 2000-350702/30.

Matches 13; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 4 PTLRQWLARAGGGGGGIEG 25
|:| | | | | | | | |
Db 10 PSLSLRLRERAGGGGGGGGAG 31

RESULT 30

O75182
ID 075182 PRELIMINARY; PRT; 1130 AA.
AC O75182;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0700 PROTEIN (FRAGMENT).
GN KIAA0700.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-98403880; PubMed-9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL; AB014600; BAA31675.1; -.
DR InterPro; IPR003822; PAH.
DR Pfam; PF02671; PAH; 3.
FT NON_TER 1
SQ SEQUENCE 1130 AA; 129359 MW; B767339317ECC96D CRC64;

Query Match 30.1%; Score 58; DB 4; Length 1130;

Best Local Similarity 54.2%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 12 ARAGGGKGGGGIEGPTLRQWLAR 35
| | | | | | | | | | | | | | |
Db 2 AHAGGGSGGGGAGGAGGLSGAR 25

Search completed: October 9, 2002, 09:03:13
Job time : 14.9826 secs

Query Match 30.1%; Score 58; DB 4; Length 524;
Best Local Similarity 53.8%; Pred. No. 57;
Matches 14; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGGKGGGIEGPTL 28
DB 377 GPLDLSALACNGGGGGGMPGL 402

RESULT 28

Q9ASE5 PRELIMINARY; PRT; 529 AA.
ID Q9ASE5
AC Q9ASE5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE P0456F08.14 PROTEIN.
GN P0456F08.14
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0456F08.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002901; BAB39414.1;
DR InterPro; IPR002937; Amino_oxidase.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF01593; Amino_oxidase; 1.
SQ SEQUENCE 529 AA; 55981 MW; 0A5DA55CDD076D24 CRC64;

Query Match 30.1%; Score 58; DB 10; Length 529;
Best Local Similarity 63.2%; Pred. No. 58;
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 LRQWLAARAGGKGGGIE 24
DB 151 LRAYQARSAGGGGGGKE 169

RESULT 29

Q9P270 PRELIMINARY; PRT; 612 AA.
ID Q9P270
AC Q9P270;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE KIAA1458 PROTEIN (FRAGMENT).
GN KIAA1458.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
genes.XVII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040891; BAA95982.1;
FT NON_TER 1
SQ SEQUENCE 612 AA; 65593 MW; 9AA4061D21E1E9FD CRC64;

Query Match 30.1%; Score 58; DB 4; Length 612;
Best Local Similarity 59.1%; Pred. No. 67;

Query Match 30.1%; Score 58; DB 11; Length 492;
Best Local Similarity 70.6%; Pred. No. 54;
Matches 12; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 LRQWLAARAGGKGGG 22
DB 384 LRQGLKTDAGGGAGGG 400

RESULT 27

Q9B2E0 PRELIMINARY; PRT; 524 AA.
ID Q9B2E0
AC Q9B2E0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KRUPPEL-LIKE ZINC FINGER PROTEIN GLIS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Zhang F., Kurebayashi S., Jettan A.M.;
RT "Cloning and genomic structure of GLIS2, a novel gene encoding a Gli-
related, Kruppel-like protein.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325914; AAK00954.1;
DR HSP; P08151; 2GLI.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 524 AA; 55704 MW; 3E2C27243DE5A85E CRC64;

Query Match 30.1%; Score 58; DB 11; Length 492;
Best Local Similarity 70.6%; Pred. No. 54;
Matches 12; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 LRQWLAARAGGKGGG 22
DB 384 LRQGLKTDAGGGAGGG 400

RESULT 27

Q9B2E0 PRELIMINARY; PRT; 524 AA.
ID Q9B2E0
AC Q9B2E0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KRUPPEL-LIKE ZINC FINGER PROTEIN GLIS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Zhang F., Kurebayashi S., Jettan A.M.;
RT "Cloning and genomic structure of GLIS2, a novel gene encoding a Gli-
related, Kruppel-like protein.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325914; AAK00954.1;
DR HSP; P08151; 2GLI.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 524 AA; 55704 MW; 3E2C27243DE5A85E CRC64;

Query Match 30.1%; Score 58; DB 11; Length 492;
Best Local Similarity 70.6%; Pred. No. 54;
Matches 12; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Wed Oct 9 10:30:07 2002

```

Query Match      30.6%; Score 59; DB 5; Length 654;
Best Local Similarity 48.0%; Pred. No. 55;
Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

Qy 6 LRQWLAARAGGKG--GGGIEGPTL 28
Db 15 LLHWASAGAGGGAGGAGLGGPAV 39

RESULT 19
Q9UAE7 PRELIMINARY: PRT; 654 AA.
ID Q9UAE7
AC Q9UAE7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROHORMONE AND NEUROPEPTIDE PROCESSING PROTEASE.
GN AMON OR PC2 OR CG6438.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99365357; PubMed=10436051;
RA Siekhaus D.E., Fuller R.S.;
RT "A role for amontillado, the Drosophila homolog of the neuropeptide
RT precursor processing protease PC2, in triggering hatching behavior.";
RL J. Neurosci. 19:6942-6954(1999).
DR EMBL: AF033117; AAD49105.1; -.
DR HSSP: P04072; ITHM.
DR MEROPS: S08.073; -.
DR FlyBase: FBgn0023179; amon.
DR InterPro: IPR000209; Peptidase_s8.
DR InterPro: IPR002884; P_domain.
DR Pfam: PF01483; P; 1.
DR Pfam: PF00082; Peptidase_S8.
DR PRINTS: PR00723; SUBTILISIN.
DR PRODOM: PD000717; P_domain; 1.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Protease; Neuropeptide.
SQ SEQUENCE 654 AA; 71733 MW; D021D4882293C996 CRC64;

Query Match      30.6%; Score 59; DB 5; Length 654;
Best Local Similarity 48.0%; Pred. No. 55;
Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

Qy 6 LRQWLAARAGGKG--GGGIEGPTL 28
Db 15 LLHWASAGAGGGAGGAGLGGPAV 39

RESULT 20
O83436 PRELIMINARY: PRT; 683 AA.
ID O83436
AC O83436;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0421.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98333770; PubMed=9665876;

Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
Khaliq H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
Venter J.C.;
"Complete genome sequence of Treponema pallidum, the syphilis
agent";
RT spirochete;
RL Science 281:375-388(1998);
DR EMBL: AF001220; AAC65409.1; -.
DR TIGR: TP0421;
DR InterPro: IPR001258; NHL.
DR InterPro: IPR001440; TPR.
DR Pfam: PF01436; NHL; 4.
DR Pfam: PF00515; TPR; 1.
KW Complete proteome.
SQ SEQUENCE 683 AA; 74518 MW; F91407FA7094AAD1 CRC64;

Query Match      30.6%; Score 59; DB 16; Length 683;
Best Local Similarity 43.8%; Pred. No. 57;
Matches 14; Conservative 2; Mismatches 12; Indels 4; Gaps 1;

Qy 4 PTLROWLAARAGGKGGGGIEGPTLROWLAAR 35
Db 74 PLILEWL-----GNAYRSYGIEGAALHGWGAAR 101

RESULT 21
Q9FT25 PRELIMINARY: PRT; 202 AA.
ID Q9FT25
AC Q9FT25;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE P0436E04.1 PROTEIN.
GN P0436E04.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0436E04.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002818; BAB16319.1; -.
SQ SEQUENCE 202 AA; 19763 MW; BFC2520037F8E274 CRC64;

Query Match      30.3%; Score 58.5; DB 10; Length 202;
Best Local Similarity 36.6%; Pred. No. 19;
Matches 15; Conservative 5; Mismatches 14; Indels 7; Gaps 1;

Qy 1 IEGPTLROWLAARAGGKG-----GGIEGPTLROWLAA 34
Db 94 VVSPSCRRQTAGRHGGCGGGRWMAAAGGGRGGCRRWMAA 134

RESULT 22
O33230 PRELIMINARY: PRT; 495 AA.
ID O33230
AC O33230;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 53.3 KDA PROTEIN.
GN HFLX OR RV2725C OR MTCY154.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```

```

RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Hopwood D.A.;
RT Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL022268; CAA18333.1; -.
KW Hypothetical protein.
SQ SEQUENCE 497 AA; 54506 MW; FB3646D4D6E54028 CRC64;

Query Match 30.6%; Score 59; DB 2; Length 497;
Best Local Similarity 51.9%; Pred. No. 42;
Matches 14; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

QY 4 PTLRW---LAARAGGKGGG-GIEGP 26
   ||| | : : ||||| ||
DB 206 PRLRGWGQSLRQMGKGKGLATGCP 232

RESULT 17
O43267 PRELIMINARY; PRT; 498 AA.
AC O43267;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE ASTROCYTE NOVAL-LIKE RNA-BINDING PROTEIN (NEUROONCOLOGIC VENTRAL
DE ANTIGEN 3) (FRAGMENT).
OS ANOVA OR NOVA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20197319; PubMed=10735272;
RA Ueki K., Ramaswamy S., Billings S.J., Mohrenweiser H.W., Louis D.N.;
RT "ANOVA, a putative astrocytic RNA binding protein gene that maps to
RT chromosome 19q13.3".
RL Neurogenetics 1:31-36(1997).
CC -!- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC
CC SUBSET OF DEVELOPING NEURONS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: BRAIN. EXPRESSION RESTRICTED TO ASTROCYTES.
CC -!- DISEASE: MAY BE A TARGET ANTIGEN IN ONE OF THE UNDEFINED HUMAN
CC PARANEOPLASTIC SYNDROMES.
DR EMBL; U70477; AAB88661.1; -.
DR InterPro; IPR004087; KH.
DR Pfam; PF00013; KH-domain; 3.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS00084; KH_TYPE_1; 3.
KW Nuclear protein; RNA-binding; Repeat; Antigen.
FT NON_TER 1
FT DOMAIN 42 76 KH.
FT DOMAIN 140 174 KH.
FT DOMAIN 246 255 POLY-ALA.
FT DOMAIN 325 330 POLY-ALA.
FT DOMAIN 350 356 POLY-PRO.
FT DOMAIN 363 368 POLY-ALA.
FT DOMAIN 375 385 POLY-GLY.
FT DOMAIN 389 397 POLY-ALA.
FT DOMAIN 416 446 KH.
SQ SEQUENCE 498 AA; 49721 MW; C4B54196FDB6BF78 CRC64;

Query Match 30.6%; Score 59; DB 4; Length 498;
Best Local Similarity 53.6%; Pred. No. 42;
Matches 15; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 9 WLAARAGGKGGGIEGPTLQWLAAARA 36
   : | | | | | | | | | | | | | | | | | |

```

```

DB 371 YLGAGAGGAGGGG--GPLVAAAAAAGA 396

RESULT 18
Q9VBC7 PRELIMINARY; PRT; 654 AA.
AC Q9VBC7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AMON PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Balslev E.M.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam K.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mergulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003757; AAF56615.1; -.
DR HSSP; P04072; 1THM.
DR MEROPS; S08.073; -.
DR FlyBase; FBgn023179; amon.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR002884; P_domain.
DR Pfam; PF01483; P; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 654 AA; 71733 MW; 2BD57F683929D237 CRC64;

```

```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Howland T.J., Wei M.-H., Ibegwam C.,
RA Hostin D., Houston K.A., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Kravitz S., Kulp D., Liang Y., Lin X.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Liang Y., Lin X.,
RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
RL EMBL; AF003474; AAF47627.1; -.
DR FlyBase; FBgn0035323; CG13807.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHLL.
SQ SEQUENCE 170 AA; 19099 MW; 477D79D55ADF4CE5 CRC64;

Query Match 30.6%; Score 59; DB 5; Length 170;
Best Local Similarity 45.8%; Pred. No. 14;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 2 EGPTLRQLAARAGGKGKGTTG 25
Db 47 EPIPVENWM-----GGGGGGGFGQ 66

RESULT 14
Q9UEA1 PRELIMINARY; PRT; 464 AA.
ID Q9UEA1
AC Q9UEA1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA-BINDING PROTEIN NOVA-2 (FRAGMENT)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99007301; PubMed=9789075;
RA Yang Y.Y., Yin G.L., Darnell R.B.;
RT "The neuronal G.L.-binding protein Nova-2 is implicated as the
RT autoantigen targeted in POMA patients with dementia.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:13254-13259(1998).
DR EMBL; AF083898; AAC72355.1; -.
DR InterPro; IPR004087; KH.
DR Pfam; PF00013; KH-domain; 3.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS50084; KH_TYPE_1; 3.
DR SEQUENCE 492 AA; 49008 MW; 41B63EAF6899256B CRC64;

Query Match 30.6%; Score 59; DB 4; Length 492;
Best Local Similarity 53.6%; Pred. No. 41;
Matches 15; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 9 WLAARAGGKGKGIEGPTLRQLAARA 36
Db 365 YLGAGAGGAGGGG--GPLVAAAAAAGA 390

RESULT 16
O69972 PRELIMINARY; PRT; 497 AA.
ID O69972
AC O69972
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOHETICAL 54.5 KDA PROTEIN.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```


RT OSJNBa0068A07, from Chromosome 10, complete sequence. *;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC091734; AAK98762.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 113 AA; 12154 MW; C1E6E87E9B00D577 CRC64;

Query Match 31.1%; Score 60; DB 10; Length 113;
 Best Local Similarity 56.5%; Pred. No. 7.1;
 Matches 13; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 14 AGGGGGGGGPTLRQWLAARA 36
 :|||:||||| :|||:|||||
 Db 26 SGGGEGGG--GRRQQRMAARA 46

RESULT 10
 Q9LWC8 PRELIMINARY; PRT; 125 AA.
 AC Q9LWC8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoideae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone: P0483F08.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002094; BAA96216.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 125 AA; 13396 MW; C609DBD0B07BC505 CRC64;

Query Match 31.1%; Score 60; DB 10; Length 125;
 Best Local Similarity 40.5%; Pred. No. 7.9;
 Matches 17; Conservative 2; Mismatches 9; Indels 14; Gaps 2;

QY 2 EGPTLRQWLAARA-----GGGGGGGEGPTLRQ 30
 ||| ||| |||
 Db 83 EGAAR-WRAARSPARGGQRRGGGGGGGRRRRRR 123

RESULT 11
 Q9HEA4 PRELIMINARY; PRT; 776 AA.
 AC Q9HEA4;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN B11A5.200.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL451109; CAC18624.2; -;
 KW Hypothetical protein.
 SQ SEQUENCE 776 AA; 82771 MW; C9BEA870D94A37DE CRC64;

Query Match 31.1%; Score 60; DB 3; Length 776;
 Best Local Similarity 53.8%; Pred. No. 50;
 Matches 14; Conservative 3; Mismatches 5; Indels 4; Gaps 2;

QY 15 GGGGGGGI---EG-PTLRQWLAARA 36
 ||| ||||| :| | |||||
 Db 678 GGGGGGGVDDDDGDPDFAGWLAQA 703

RESULT 12
 Q9JMH4 PRELIMINARY; PRT; 1431 AA.
 ID Q9JMH4;
 AC Q9JMH4;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE COLLAGEN TYPE XVII.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;
 RT "Mesocricetus auratus mRNA for type XVII collagen.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB027759; BAA94381.1; -;
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 5.
 SQ SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9A5C CRC64;

Query Match 30.8%; Score 59.5; DB 11; Length 1431;
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 12 ARAGGGGGGIEGPTLRQWLAARA 36
 ||| ||| ||| | | | | |
 Db 438 ARGGGGGGGGGGGT--WGAAPA 459

RESULT 13
 Q9W033 PRELIMINARY; PRT; 170 AA.
 ID Q9W033;
 AC Q9W033;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG13807 PROTEIN.
 GN CG13807.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

Wed Oct 9 10:30:07 2002

```

RA "Massive gene decay in the leprosy bacillus.";
RT Nature 409:1007-1011(2001).
RL EMBL; AL583920; CAC31378.1; -.
DR InterPro; IPR000765; GTP1_OBG.
DR PRINTS; PR00326; GTP1_OBG.
KW COMPLETE PROTEOME.
SQ SEQUENCE 488 AA; 52800 MW; 188918856F9774AA CRC64;

Query Match 31.3%; Score 60.5; DB 16; Length 488;
Best Local Similarity 43.3%; Pred. No. 27;
Matches 13; Conservative 2; Mismatches 8; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGKGGGIEGP 26
DB 189 PRLRGWGESMSRQVGRAGGGVGLRGP 218

RESULT 8
Q49843 PRELIMINARY; PRT; 518 AA.
ID Q49843
AC Q49843;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HFLX.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00019; AAA17274.1; -.
SQ SEQUENCE 518 AA; 56001 MW; 6641916CC84F374B CRC64;

Query Match 31.3%; Score 60.5; DB 2; Length 518;
Best Local Similarity 43.3%; Pred. No. 29;
Matches 13; Conservative 2; Mismatches 8; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGKGGGIEGP 26
DB 219 PRLRGWGESMSRQVGRAGGGVGLRGP 248

RESULT 9
Q947T7 PRELIMINARY; PRT; 113 AA.
ID Q947T7
AC Q947T7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 12.2 KDA PROTEIN.
GN OSJNBA0068A07.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA de la Bastide M.; Spiegel L.; Kirchoff K.; Preston R.; King L.;
RC STRAIN-NIPPONBARE;
RA Vil M.D.; Baker J.; Zutavern T.; Santos L.; Miller B.; Kuit K.;
RA Cunius D.M.; Bell M.; Ballija V.; Shah R.; Bahret A.; Dike S.;
RA Yang C.; O'Shaughnessy A.; Palmer L.; Dedhia N.; McCombie W.R.;
RA "Genomic Sequence for Oryza sativa, Nipponbare strain, clone

```

Wed Oct 9 10:30:07 2002

```

Db      80 GPTVGVVAYRAGAGGGGGPRGFALK 106

RESULT 4
Q9SDK6 PRELIMINARY; PRT; 439 AA.
AC Q9SDK6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
   clone: P0705D01";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RW EMBL; AP000492; BAA84610.1;
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 47297 MW; 533EEC240CEA1BA2 CRC64;

Query Match 32.1%; Score 62; DB 10; Length 439;
Best Local Similarity 32.0%; Pred. No. 17;
Matches 16; Conservative 2; Mismatches 18; Indels 14; Gaps 1;

QY 1 IEPTTLROWLAARAGGKGGGG-----IEGPTTLROWLAARA 36
   : ||| | ||| ||| | : | | |
Db 39 LHAPLLRLPLGGGGGGGGGGGGGGERVGVAVRGVGEARSQRAEA 88

RESULT 5
Q19476 PRELIMINARY; PRT; 500 AA.
AC Q19476;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F15B9.5 PROTEIN.
GN F15B9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
   investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 278013; CAB01420.1;
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR Hydrolase; Serine protease.
SQ SEQUENCE 500 AA; 53946 MW; 1416327086FE7CF6 CRC64;

Query Match 31.6%; Score 61; DB 5; Length 500;
Best Local Similarity 52.2%; Pred. No. 25;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 GPTTLROWLAARAGGKGGGGIEG 25
   : ||| | ||| | ||| |
Db 429 GSMGLRFLNRGGGGGGGGGGG 451

Q9PVG9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POU-BOX PROTEIN BRAIN-2.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Xue J.X., Zhang W., Fu D.C., He R.Q., Xue Z.G.;
RT "grain-2, a POU-box gene expressed in quail embryos.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: WITH OTHER HOMEOBOX PROTEINS.
DR EMBL; AF091043; AAF00040.1;
DR HSP; P14859; LOC.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000327; POU.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PF00157; pou; 1.
DR PRODOM; PD000583; POU; 1.
DR SMART; SM00352; POU; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS50071; HOMEOBOX_2; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 431 AA; 43722 MW; 1DC47E53F9ACCT7D5 CRC64;

Query Match 33.9%; Score 65.5; DB 13; Length 431;
Best Local Similarity 40.5%; Pred. No. 6.4;
Matches 17; Conservative 2; Mismatches 6; Indels 17; Gaps 2;

QY 8 OWLAARA-----GGKGGGIEGPTTLROWLAARA 36
   ||| | : ||| ||| | | | | |
Db 58 QWTAALSHGGGGGGGGGGGGGGGEAP-----WAAAAA 95

RESULT 3
Q943K0 PRELIMINARY; PRT; 253 AA.
AC Q943K0;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE P0039A07.6 PROTEIN.
GN P0039A07.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
   clone: P0039A07";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RW EMBL; AP003235; BAB64100.1;
SQ SEQUENCE 253 AA; 25568 MW; A963166CE5F97B2B CRC64;

Query Match 32.6%; Score 63; DB 10; Length 253;
Best Local Similarity 51.9%; Pred. No. 7.3;
Matches 14; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 3 GPTTLROWLAARAGGKGGGGIEGPTTLR 29
   ||| | : ||| | ||| | | |

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-30

Perfect score: 193

Sequence: 1 IEPTLRQWLARAGGKGGGIEGPTLRQWLARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	34.2	360	10 Q9LGC9	Q9lgc9 oryza sativ
2	65.5	33.9	431	13 Q9PVG9	Q9pv9 coturnix co
3	63	32.6	253	10 Q943K0	Q943k0 oryza sativ
4	62	32.1	439	10 Q9SDK6	Q9sd6 oryza sativ
5	61	31.6	500	5 Q19476	Q19476 caenorhabdi
6	60.5	31.3	407	2 Q9L0B6	Q9l0b6 streptomyce
7	60.5	31.3	488	16 Q9CCCO	Q9ccc0 mycobacteri
8	60.5	31.3	518	2 Q9843	Q9843 mycobacteri
9	60	31.1	113	10 Q947T7	Q947t7 oryza sativ
10	60	31.1	125	3 Q9LWC8	Q9lwc8 oryza sativ
11	60	31.1	776	3 Q9HEA4	Q9hea4 mesocricetu
12	59.5	30.8	1431	11 Q9JMH4	Q9jmh4 drosophila
13	59	30.6	170	5 Q9W033	Q9w033 drosophila
14	59	30.6	464	4 Q9UEA1	Q9uea1 homo sapien
15	59	30.6	492	4 Q9UNW9	Q9unw9 homo sapien
16	59	30.6	497	2 Q69572	Q69572 streptomyce

17	59	30.6	498	4	043267	043267 homo sapien
18	59	30.6	654	5	Q9VBC7	Q9vbc7 drosophila
19	59	30.6	654	5	Q9UAE7	Q9uae7 drosophila
20	59	30.6	683	16	083436	083436 treponema p
21	58.5	30.3	202	10	Q9FTZ5	Q9ftz5 oryza sativ
22	58.5	30.3	495	16	033230	033230 mycobacteri
23	58.5	30.3	496	2	Q9AD76	Q9ad76 streptomyce
24	58	30.1	156	10	Q94EA0	Q94ea0 oryza sativ
25	58	30.1	377	13	Q9VHD0	Q9vhd0 petromyzon
26	58	30.1	492	11	035392	035392 mus musculu
27	58	30.1	524	4	Q9BZE0	Q9bze0 homo sapien
28	58	30.1	529	10	Q9ASE5	Q9ase5 oryza sativ
29	58	30.1	612	4	Q9P270	Q9p270 homo sapien
30	58	30.1	1130	4	075182	075182 homo sapien
31	57.5	29.8	243	10	Q9AR44	Q9ar44 oryza sativ
32	57.5	29.8	392	10	Q9ZRB9	Q9zrb9 lycopersico
33	57.5	29.8	1548	4	Q9NVW9	Q9nvw9 homo sapien
34	57.5	29.8	2161	4	Q9Y566	Q9y566 homo sapien
35	57	29.5	447	13	073628	073628 anolis caro
36	57	29.5	452	5	Q9VJK4	Q9vjk4 drosophila
37	57	29.5	491	10	Q94LP1	Q94lp1 oryza sativ
38	57	29.5	607	2	Q9L8D4	Q9l8d4 polyangium
39	56.5	29.3	244	11	Q9D384	Q9d384 mus musculu
40	56.5	29.3	302	2	Q9S596	Q9s596 myxococcus
41	56	29.0	103	10	Q9M045	Q9m045 arabidopsis
42	56	29.0	249	10	049673	049673 arabidopsis
43	56	29.0	309	5	Q9VV01	Q9vv01 drosophila
44	56	29.0	331	5	Q9U2I1	Q9u2i1 caenorhabdi
45	56	29.0	333	5	Q9U2I0	Q9u2i0 caenorhabdi

ALIGNMENTS

RESULT 1
Q9LGC9 PRELIMINARY; PRT; 360 AA.
ID Q9LGC9
AC Q9LGC9;
DC 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RT Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0462H08.19";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002525; BAB07996.1;
DR InterPro; IPR000571; zf-CCCH.
DR Pfam; PF00642; zf-CCCH; 4.
DR SMART; SM00356; Znf_C3H1; 4.
SQ SEQUENCE 360 AA; 37368 MW; 5105598B7E1C77B2 CRC64;
Query Match 34.2%; Score 66; DB 10; Length 360;
Best Local Similarity 52.0%; Pred. No. 4.7;
Matches 13; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
Oy 1 IEPTLRQWLARAGGKGGGIEG 25
:|||||:|||||:
Db 26 LEGPWRMGLGGGGGGGGGG 50
RESULT 2
Q9PVG9 PRELIMINARY; PRT; 431 AA.
ID Q9PVG9

```
AC Q943K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P0039A07.6 PROTEIN.
GN P0039A07.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta.
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0039A07."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003235; BAB64100.1; -.
SQ SEQUENCE 253 AA; 25568 MW; A963166CE5F97B2B CRC64;

Query Match 37.4%; Score 71; DB 10; Length 253;
Best Local Similarity 55.6%; Pred. No. 0.86;
Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 3 GPTLRQCLARAGGGGGGIEGPTLR 29
|||: | ||| ||||| | |
Db 80 GPTGVRYVAYRAGAGGGGGPRGFALK 106

RESULT 3
Q9LYB2
ID Q9LYB2 PRELIMINARY; PRT; 199 AA.
AC Q9LYB2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 21.5 KDA PROTEIN.
GN T20010.200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Ottenwaelder B., Duchemin D., Zeitler K., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163816; CAB87755.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR002395; Kininogen.
DR PRINTS; PR00334; KINNOGEN.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 199 AA; 21539 MW; E5D28AC167B3FBF8 CRC64;

Query Match 35.8%; Score 68; DB 10; Length 199;
Best Local Similarity 34.8%; Pred. No. 1.5;
Matches 16; Conservative 3; Mismatches 11; Indels 16; Gaps 1;

QY 2 EGPTLRQC-----LAARAGGGGGGIEGPTLRQC 31
|||: | ||| ||||| | |
Db 7 EGRTRKCPFASTCTSLVAQTSLLCVDGGGGGGVGVDRGC 52

RESULT 4
Q9P270
ID Q9P270 PRELIMINARY; PRT; 612 AA.
```

```
AC Q9P270;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA1458 PROTEIN (FRAGMENT).
GN KIAA1458.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain
RL which code for large proteins in vitro."
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040891; BAA95982.1; -.
FT NON_TER 1
SQ SEQUENCE 612 AA; 65593 MW; 9AA4061D21E1E9FD CRC64;

Query Match 35.8%; Score 68; DB 4; Length 612;
Best Local Similarity 63.6%; Pred. No. 4.4;
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLRQCLARAGGGGGGIEG 25
|||: | ||||| ||| |
Db 10 PSLSLRLRERAGGGGGGGAG 31

RESULT 5
Q91BC5
ID Q91BC5 PRELIMINARY; PRT; 66 AA.
AC Q91BC5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 7.0 KDA PROTEIN.
OS Spodoptera litura nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G2;
RX MEDLINE=21425398; PubMed=11531416;
RA Pang Y., Yu J., Wang L., Hu X., Bao W., Li G., Chen C., Han H., Hu S.,
RA Yang H.;
RT "Sequence Analysis of the Spodoptera litura Multicapsid
RT Nucleopolyhedrovirus Genome."
RL Virology 287:391-404(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=G2;
RA Yu J., Wang L., Hu X., Pang Y.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325155; AAL01786.1; -.
KW Hypothetical protein.
SQ SEQUENCE 66 AA; 6998 MW; C5626A8FFA9C9E7C CRC64;

Query Match 34.7%; Score 66; DB 12; Length 66;
Best Local Similarity 54.5%; Pred. No. 0.89;
Matches 12; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 7 RQCLARAGGGGGGIEGPTL 28
|||: | ||||| ||| |
Db 13 QQASSNRSGGGGGGVGAML 34

RESULT 6
Q9M6A1
ID Q9M6A1 PRELIMINARY; PRT; 137 AA.
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 : Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-27
Perfect score: 190
Sequence: 1 IEPTLRQCLAAAGGGGGGIEPTLRQCLAAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	38.4	360	10 Q9LGC9	Q9LGC9 oryza sativ
2	71	37.4	253	10 Q943K0	Q943K0 oryza sativ
3	68	35.8	199	10 Q9L1B2	Q9L1B2 arabidopsis
4	68	35.8	612	4 Q9P270	Q9P270 homo sapien
5	66	34.7	66	12 Q91BC5	Q91BC5 spodoptera
6	66	34.7	137	10 Q9M6A1	Q9M6A1 catharanthu
7	66	34.7	160	10 Q9M699	Q9M699 catharanthu
8	66	34.7	369	10 Q9XE89	Q9XE89 sorghum bic
9	66	34.7	413	10 Q9L1B6	Q9L1B6 oryza sativ
10	66	34.7	474	4 Q96SQ2	Q96SQ2 homo sapien
11	66	34.7	496	2 Q9AD76	Q9AD76 streptomyce
12	66	34.7	500	5 Q19476	Q19476 caenorhabdi
13	66	34.7	688	4 Q9BYD8	Q9BYD8 homo sapien
14	66	34.7	689	4 Q96JG7	Q96JG7 homo sapien
15	66	34.7	707	11 Q61869	Q61869 mus musculu
16	66	34.7	752	4 Q96134	Q96134 homo sapien

17 65.5 34.5 355 4 Q9Y648 Q9Y648 homo sapien
18 65.5 34.5 770 5 Q9GNP1 Q9GNP1 ciona savi
19 65 34.2 381 10 Q9LD54 Q9LD54 oryza sativ
20 65 34.2 452 5 Q9VJK4 Q9VJK4 drosophila
21 64.5 33.9 150 5 Q9VP99 Q9VP99 drosophila
22 64.5 33.9 165 2 Q9AF15 Q9AF15 mycobacteri
23 64 33.7 309 5 Q9VW01 Q9VW01 drosophila
24 64 33.7 331 5 Q9U2I1 Q9U2I1 caenorhabdi
25 64 33.7 333 5 Q9U2I0 Q9U2I0 caenorhabdi
26 64 33.7 422 5 Q96755 Q96755 branchiost
27 64 33.7 529 10 Q9ASE5 Q9ASE5 oryza sativ
28 64 33.7 3626 2 Q9F7T9 Q9F7T9 streptomyce
29 64 33.7 3972 2 Q9S0R8 Q9S0R8 streptomyce
30 63.5 33.4 113 10 Q942U6 Q942U6 oryza sativ
31 63.5 33.4 434 10 Q9M1I8 Q9M1I8 arabidopsis
32 63 33.2 146 3 Q9C1E7 Q9C1E7 schizophyll
33 63 33.2 186 10 Q942R8 Q942R8 oryza sativ
34 63 33.2 488 16 Q9CCCC Q9CCCC mycobacteri
35 63 33.2 518 2 Q49843 Q49843 mycobacteri
36 63 33.2 796 5 Q27258 Q27258 drosophila
37 63 33.2 797 5 Q9V7U9 Q9V7U9 drosophila
38 63 33.2 806 5 Q96828 Q96828 drosophila
39 63 33.2 841 10 Q9SXI9 Q9SXI9 oryza sativ
40 63 33.2 1186 5 Q61080 Q61080 acanthamoeb
41 62.5 32.9 339 10 Q9S270 Q9S270 arabidopsis
42 62.5 32.9 775 4 Q9C0I1 Q9C0I1 homo sapien
43 62.5 32.9 867 5 Q9W149 Q9W149 drosophila
44 62 32.6 125 10 Q9LWC8 Q9LWC8 oryza sativ
45 62 32.6 168 10 Q9FTK4 Q9FTK4 oryza sativ

ALIGNMENTS

RESULT 1

ID Q9LGC9 PRELIMINARY; PRT; 360 AA.
AC Q9LGC9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0462H08.19";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002525; BAB07996.1; -;
DR InterPro; IPR000571; Zf-CCCH.
DR Pfam; PF00642; zf-CCCH; 4.
DR SMART; SM00356; Znf_C3H1; 4.
SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 38.4%; Score 73; DB 10; Length 360;
Best Local Similarity 56.0%; Pred. No. 0.72;
Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAAAGGGGGGIEPTLRQCLAAARA 36

Db 26 LEGPWRMRLGCGGGGGGGGGGG 50

RESULT 2

Q943K0 PRELIMINARY; PRT; 253 AA.
ID Q943K0